Sibling Relationships: A Systematic Review of Measures and the Development and Validation of a New Measure

Emily Ralph, BSc, MSc

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology

School of Health and Wellbeing
College of Medical, Veterinary and Life Sciences
University of Glasgow

February 2025

Declaration of Originality Form

This form **must** be completed and signed and submitted with all assignments.

Please complete the information below (using BLOCK CAPITALS).

Name EMILY RALPH

Student Number xxxxxxx

Course Name **Doctorate in Clinical Psychology**

Assignment Number/Name SIBLING RELATIONSHIPS: A SYSTEMATIC REVIEW OF MEASURES AND THE DEVELOPMENT AND VALIDATION OF A NEW MEASURE

A link to the University's Statement on Plagiarism is provided at the end of this form. Please read the Statement on Plagiarism carefully THEN read and sign the declaration below.

I confirm that this assignment is my own work and that I have:

Read and understood the guidance on plagiarism in the Doctorate in Clinical Psychology Programme Handbook, including the University of Glasgow Statement on Plagiarism

Clearly referenced, in both the text and the bibliography or references, **all sources** used in the work

Fully referenced (including page numbers) and used quotation marks for **all text quoted** from books, journals, web etc. using a consistent author-date referencing system

Provided the sources for all tables, figures, data etc. that are not my own work

Not made use of the work of any other student(s) past or present without acknowledgement. This includes any of my own work, that has been previously, or concurrently, submitted for assessment, either at this or any other educational institution, including school (see overleaf at 31.2)

Not sought or used the services of any professional agencies

Not used, without acknowledgment, any website or software (including generative artificial intelligence) that generates assessment responses

In addition, I understand that any false claim in respect of this work will result in conduct action in accordance with University regulations

DECLARATION:

I am aware of and understand the University's policy on plagiarism and I certify that this assignment is my own work**,** except where indicated by referencing, and that I have followed the good academic practices noted above

Signed EMILY RALPH

Declaration of word count for submission of DClinPsy thesis for examination

Trainee name:	EMILY RALPH
Matriculation number:	xxxxxxx
Title of thesis:	SIBLING RELATIONSHIPS: A SYSTEMATIC REVIEW OF MEASURES AND THE DEVELOPMENT AND VALIDATION OF A NEW MEASURE
Systematic Review chapter word count: *Must adhere to (a) target journal word limit or (b) 6,000-9,000 words including abstract, main text, tables, figures & references*	7,220
Major Research Project chapter word count (excluding plain language summary): *Must adhere to (a) target journal word limit or (b) 6,000-9,000 words including abstract, main text, tables, figures & references*	7,926
Thesis word count: *Maximum 30,000 words including all content except thesis title page, declaration forms, table of contents, list of tables, list of figures, acknowledgements* **This limit must not be exceeded**	18,724
Trainee signature:	EMILY RALPH
Date:	21ST FEBRUARY 2025

Version Control
v1.1 – 15 August 2022 – approved by Breda Cullen

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I thank my supervisors, Dr Andy Siddaway and Professor Hamish McLeod, for their invaluable guidance, expertise, and encouragement throughout this project. Their support has made this research possible, and I am grateful for the time and insight they have generously shared. I also thank Dr Jala Rizeq, who provided thoughtful and supportive advice as a research advisor. A thank you to all the participants who kindly gave their time to be part of this study; this work would not have been possible without you. I am also incredibly grateful to my colleagues on my clinical placements, not only for helping me to reach participants but also for their emotional and practical support as I navigated the challenges of this course.

To my parents, Bryony and Jeff Ralph, thank you for ensuring I had access to the opportunities that have shaped my academic journey. Your support has been the foundation that allowed me to reach this stage. To my sister, Ailsa, and my brother, Alex, though perhaps unintentionally, you both inspired my passion for this research area. As my siblings, you have been my longest and most formative relationships, shaping the person and clinical psychologist I am today. A special thank you to my Grandpa, Douglas Reid, for your wisdom, encouragement, and belief in me. I am also deeply grateful to the rest of my family for their ongoing love and support.

To my fiancée, Chris Parrott, thank you for standing by my side. Your patience, sacrifices, and encouragement have played a vital role in helping me to achieve this. I also thank Pauline and Gerard Parrott for their generosity and support, including opening their home to me at the start of my psychology journey.

To my friends, near and far, you have kept me grounded, sane, and smiling throughout this process. Your support, encouragement and ability to remind me not to take life too seriously have meant more than I can express. Thank you for the adventures.

Finally, to all the teachers, lecturers, supervisors, and colleagues who have helped me believe in myself. As someone with ADHD who struggled to reach their potential in school, I never imagined I would get this far. Thank you.

Chapter 1

Assessing Sibling Relationships: A Systematic Review of Measures and Their Psychometric Status

Prepared in accordance with the author requirements for Family Relations Journal; https://onlinelibrary.wiley.com/page/journal/17413729/homepage/forauthors.html

Abstract

Sibling relationships play a significant role in children's emotional and social development, and several instruments have been developed to assess this dynamic. However, it is unclear which measures are available and how many of these meet the current recommendations for robust psychometric tools. This study identified and reviewed the psychometric properties of measures of sibling relationships in children and adolescents, following the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. We also provide recommendations to guide the selection of existing measures and make suggestions for developing and validating future sibling relationship measures. We searched six electronic databases (APA PsycINFO, APA PsycTests, CINAHL, Medline, and OVID), screened studies against the inclusion and exclusion criteria, and conducted a narrative synthesis. Overall, no single measure achieved high ratings across the COSMIN criteria. Although many demonstrated moderate to good psychometric properties, for structural validity and internal consistency, each exhibited some limitations in other domains, such as content validity and cross-cultural validity. Our findings highlight the need to develop and validate a new sibling measure that aligns with modern psychometric standards. Future efforts should focus on developing and validating measures that include samples of children and/or parents across multiple developmental stages and diverse or non-traditional family structures. Additionally, comprehensive Patient-Reported Outcome Measure (PROM) development is essential to ensure strong content validity and to capture the full range of constructs relevant to sibling relationships.

Introduction

Siblings often share a bond that outlasts many other relationships, making it one of the most enduring connections in a person's life (Buhrmester & Furman, 1990; Volling, 2003). These relationships not only support emotional and social development but also facilitate the acquisition of essential interpersonal skills (Buist & Vermande, 2014; Kramer & Conger, 2009). Although siblings can display warmth and affection, their interactions may also involve conflict and aggression (Tucker & Finkelhor, 2017). Nonetheless, when these bonds are positive, they offer a source of security during challenging circumstances, buffering the impact of stressors such as poverty, martial conflict, and divorce (Milesky & Levitt, 2005; Hetherington, 1989) and promote resilience in the face of adversity (Dunn, 2006).

Despite their crucial role, sibling relationships remain relatively understudied compared to parent-child or peer relationships (McHale, Updegraff & Whiteman, 2012). Assessing sibling relationship quality is essential for both research and clinical practice. For example, clinicians may evaluate sibling relationship quality to inform family interventions, while researchers studying sibling dynamics require reliable tools to capture key relational dimensions. Several instruments have been developed to measure sibling relationships; however, a recent review (Holmes et al., 2024) highlighted that it is unclear how many of the current measures of sibling relationship quality meet the current specific criteria by structured tools for developing reliable and valid measures. They recommended that future research systematically examine each relationship quality measure to utilise robust instrument development and validation standards. There may have been a lack of rigorous development process across measures of sibling relationship quality, which could compromise their appropriateness across different subpopulations of children, limit their developmental specificity (e.g. measures for school-age siblings compared to adolescent siblings) and compromise their use of subsequent modelling. When developing and validating measures, poor construct measurement can result in unreliable or biased data, potentially compromising research findings and clinical interventions. To avoid these limitations, this systematic review applies the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) (Mokkink et al., 2024) guidelines to

identify and evaluate measures of sibling relationship quality in children and adolescents. Our findings aim to facilitate the assessment of sibling dynamics and enhance the understanding of sibling relationships in developmental contexts.

Aims

The primary aim of this systematic review is to critically evaluate the existing psychometric measures of sibling relationship quality in children and adolescents using the COSMIN guidelines to determine their methodological robustness and utility in both research and clinical contexts. The secondary aim is to guide the selection of appropriate tools and inform future development and validation of sibling relationship measures.

Objectives

1. Identify all existing measures that assess sibling relationship quality in children and adolescents up to 18 years.
2. Describe these measures' development and psychometric properties, including validity, reliability and responsiveness.
3. Evaluate the methodological quality of included studies using the COSMIN Risk of Bias checklist.
4. Provide recommendations to help clinicians and researchers select the most suitable measures based on psychometric strength and methodological rigour.
5. Suggest directions for future research in developing and validating sibling relationship measures.

Methods

Search Strategy

This review was conducted by a research team of two professors (HM, AD) and two doctoral students (ER, AD) at the University of Glasgow. A research librarian was consulted on developing and conducting an electronic search strategy to identify sibling relationship quality measures. The search was performed across six electronic databases, OVID (EMBASE, Medline, Health and Psychosocial Instruments) and

EBSCOhost (APA PsycINFO, APA PsycTests, CINAHL) on October 9, 2024. Search terms were tailored to meet the specific syntax requirements of each database. Key search terms included combinations of “sibling relationship*,” “sibling bond*,” “sibling quality,” and related terms, alongside “measure*,” “tool*,” “questionnaire*,” and “psychometric*,” including specific psychometric properties such as “validity,” “reliability,” and “factor analysis.” Full details of the search strategy are provided in Appendix 1.1.

Selection Criteria

To be included in this review, articles needed to present psychometric evidence regarding measures of sibling relationships in children and adolescents up to 18 years. Eligible studies reported psychometric evidence and could involve developing new measures or validating existing tools. Only studies published in peer-reviewed journals and available in English were included. Studies were excluded if they did not report psychometric evidence, such as reliability or validity data, focused on sibling relationships without reporting psychometric properties of the measures used, or exclusively assessed twin relationships without extending the tool for non-twin siblings. Grey literature, such as dissertations, theses, conference abstracts, reports, or book chapters, were also excluded.

Data Extraction

Articles were entered into the Rayyan online reviewing system (https://rayyan.ai) for study authors to determine whether studies met the inclusion criteria for this review. The primary researcher (ER) searched, screened, and extracted articles, and reference lists of selected articles were hand-searched for additional relevant studies. A co-rater (AD) reviewed 20% (*N* = 8) of the randomly chosen articles against the eligibility criteria at title/abstract screening and full-text screening to increase reliability. A data extraction tool (Appendix 1.2) was developed to extrapolate information relating to authors, publication year, study design, sample characteristics, recruitment, and outcome measures and to illustrate key findings from all included studies.

Narrative Synthesis

This review followed Popay et al.'s (2006) narrative synthesis framework. First, articles were described and grouped. Next, we explored potential sources of heterogeneity, such as measurement format (child vs. parent report), sample age range, cultural context and constructs measured. Finally, we assessed the robustness of our synthesis by mapping findings onto the COSMIN quality appraisal, considering methodological strengths and gaps. The results were discussed in relation to the review's aims.

Assessment of Measurement Properties: Methodological Quality

The quality appraisal tool used was the COSMIN checklist (Mokkink, 2024). The COSMIN contains questions on patient-reported outcome development (PROM), content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness (Appendix 1.3). Studies were rated as *Very Good*, *Adequate*, *Doubtful* or *Inadequate*. To ensure inter-rater reliability, 20% (*N* = 3) of the papers were randomly assigned to a second rater (AD) for independent appraisal. Discrepancies in ratings were resolved through discussion.

Cohen's kappa was calculated to assess agreement beyond chance k = 0.71 (SE = 0.11, 95% CI [0.50, 0.92]. A weighted Kappa was computed using a linear weighting scheme to account for the degree of disagreement, weighted *k* = 0.73, similarly indicating substantial agreement. This suggests a high level of agreement between raters (Landis & Koch, 1977)

Results

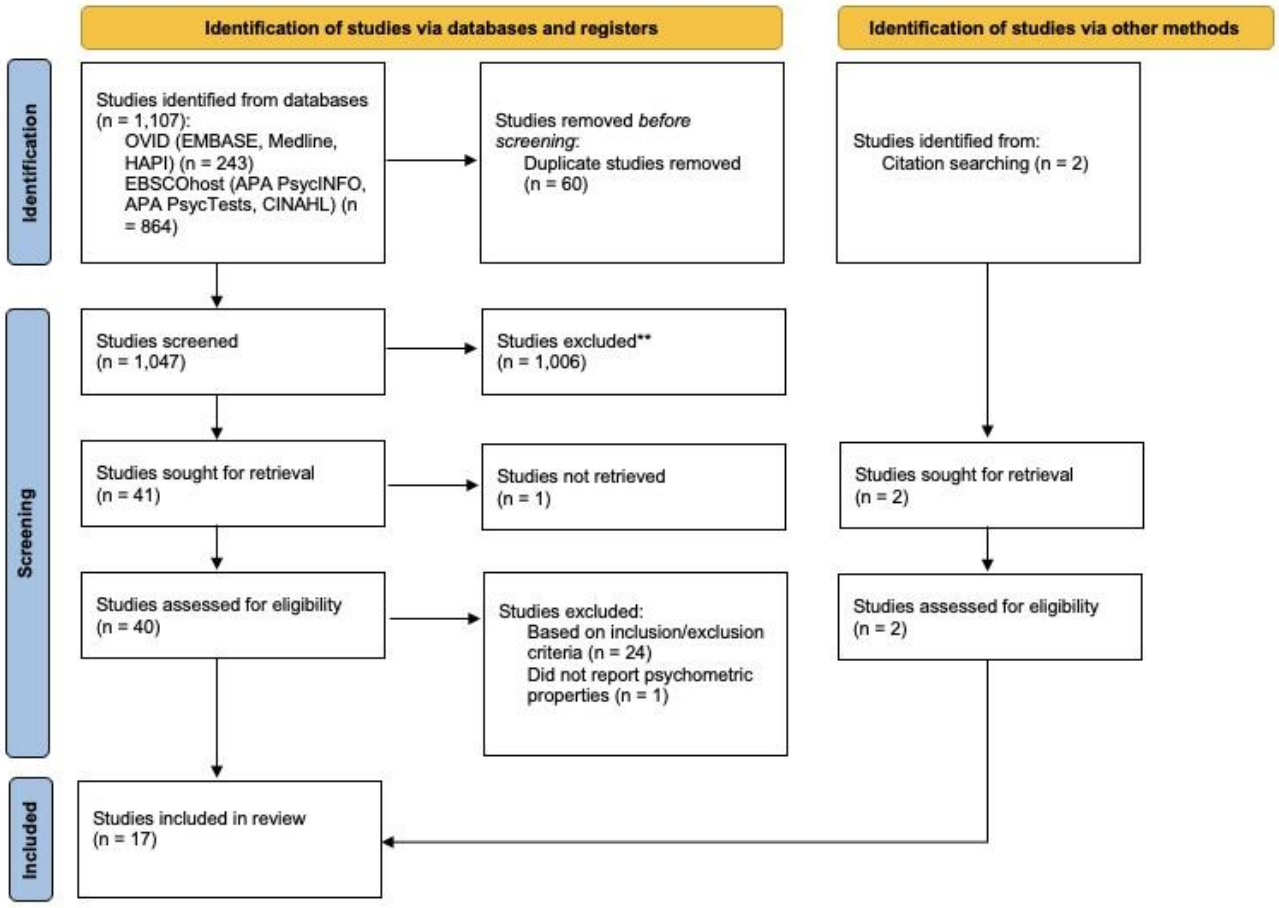
Study Selection

The search strategy yielded 1,107 records. After the removal of 60 duplicates, 1,047 abstracts were screened. Of these, 1,006 were excluded based on title and abstract review, and one full text could not be retrieved, leaving 40 full texts for full-text screening. Twenty-four articles were excluded for not meeting the eligibility criteria, and one was excluded for not reporting the psychometric properties of the measure

(Figure 1.1). Two further studies were identified through citation searching and were eligible for inclusion. Seventeen studies met the inclusion criteria and were included in this systematic review.

Figure 1.1

PRISMA Flow Diagram



Study Characteristics

Table 1.1 provides an overview of the synthesised key information from the included studies. Nine studies focused on the development of a new sibling relationship measure, while eight investigated the validation of existing instruments. Nine studies were conducted in the United States, three in China, two in Canada, two in the Netherlands, one in Spain, one in Italy, and one in Greece. The samples included 3,227

children and 2,569 parents (*n*= 5,796), with children's ages ranging from 1 month to 18 years. Most samples were predominantly White, although one study was described as "ethnically diverse" (Persram et al., 2022) without further detail. Some studies recruited international samples (Greek, Dutch, Spanish, and Chinese). Families were generally middle- or upper-middle-class, often representing intact, two-parent households. One exception included both adoptive and non-adoptive two-parent families (Stocker et al., 1989). Data collection methods included school-based, home-based, and research-based interviews, observations, and questionnaires.

Included Instruments

Ten sibling relationship measures were identified: the Sibling Relationship Questionnaire (SRQ) and four validation studies, Maternal Interview (MIACSR), Sibling Relationship Inventory (SRI) and two validation studies, Parental Expectations and Perceptions of Children's Sibling Relationship Questionnaire (PEPC-SRQ) and one validation study, Sibling Attachment Inventory (SAI), Adolescents' Perceptions of Sibling Trust (APST), Sibling Relationship Questionnaire for Chinese Preschool Children–Parental Version (SRQ-CPC-PV), Early Childhood–Sibling Relationship Questionnaire–China Specific (EC-SRQ-CS), Sibling Relationship Assessment (SRA), and Chinese–Sibling Behaviour Inventory (C-SIB).

Among the included instruments, the SRQ had the most validation studies (*n*=4), followed by the SRI (*n*=2). The MIACSR, PEPC-SRQ, and C-SIB each had one associated validation study. The remaining five measures (SAI, APST, PEP-SRQ-PV, EC-SRQ, and SRA) were included only in their original development or application reports, with no additional independent validation studies identified.

Table 1.1

Study Characteristics

Authors	Year	Journal	Country	Study Design Label	Measure	Sample Size	Age Range	Population	Setting
Furman & Buhrmester	1985	Child Development	United States	Instrument Development	Sibling Relationship Questionnaire (SRQ)	247 children	11-13 years old	Mostly Caucasian children from middle to upper middle-class families	School-based interviews and questionnaire administration
Stocker, Dunn & Plomin	1989	Child Development	United States	Instrument Development	Maternal Interview About Children's Sibling Relationships	96 mothers 192 children	3-10 years old	Intact Caucasian families, mix of adoptive families and non-adoptive families	Home-based observations and interviews
Buhrmester & Furman	1990	Child Development	United States	Instrument Validation	Sibling Relationship Questionnaire (SRQ)	363 children	8-18 years old	Predominantly Caucasian children from middle and upper-middle class families	School-based questionnaire administration
Stocker & McHale	1992	Journal of Social and Personal Relationships	United States	Instrument Development	Sibling Relationship Inventory (SRI)	206 children	6-11 years old	Children from middle class families, assumed to be Caucasian	Home-based interviews and telephone interviews
Karmer & Baron	1995	Family Relations	United States	Instrument Development	Parental Expectations and Perceptions of Children's Sibling Relationships Questionnaire (PEPC-SRQ)	114 parents 57 children	14 months-5 years old	98% White, educated, middle class two-parent families	Parents completed the questionnaire at home

Boer, Westenberg, McHale, Updegraff & Stocker	1997	Journal of Social and Personal Relationships	United States & Netherlands	Instrument Validation	Sibling Relationship Inventory (SRI)	206 American children 452 Dutch children	5-12 years old	Predominantly White, middle and working-class children	Interviews
Swift,Taylor, Kaugars, Drotar, Yeates, Wade & Stancin	2003	Developmental and Behavioual Paediatrics	United States	Instrument Validation	Sibling Relationship Questionnaire (SRQ)	103 children who were siblings of children with TBI or orthopaedic injuries	6-12 years old	Middle to upper income children with a sibling having moderate or severe TBI or orthopaedic injuries	Questionnaire-based assessments, mailed surveys, and follow-up clinic visits
Lecce, de Bernart, Vezzani, Pinto & Primi	2010	European Journal of Developmental Psychology	Italy	Instrument Validation	Sibling Relationship Inventory (SRI)	385 children	6-12 years old	Middle class children, presumed to be White	School-based assessments
Derkman, Scholte, Van der Veld & Engels	2010	European Journal of Assessment	Netherlands	Instrument Validation	Sibling Relationship Questionnaire (SRQ)	428 adolescents	13-16 years old	96-98% Dutch adolescents with biologically related siblings	Home-based data collection via interviews and questionnaires
Adamis, Tsamparl & Talanti	2017	Psychology: The Journal of Hellenic Psychological Society	Greece	Instrument Validation	Sibling Relationship Questionnaire (SRQ)	185 adolescents	10-18 years old	Adolescents in Greek Urban and rural areas	Schools for adolescent participants and homes for parent participants
Noel, Francies & Tilley	2018	Child Psychiatry and Human Development	Canada	Instrument Development	Sibling Attachment Inventory (SAI)	172 children and adolescents	10-14 years old	98% White, 77% two parent home children and adolescents	School-based data collection during class hours

Wang, Li, Liu, Zhao, Li & Niu	2021	Family Relations	China	Instrument Development	Early Childhood Sibling Relationship Questionnaire (China-specific)	954 parents of children	Parents of 1 month to 8-year-old children	Educated, Chinese parents of preschool children	Interviews and questionnaires
Persram, Howe & Bukowski	2022	Canadian Journal of Behavioural Science	Canada	Instrument Development	Adolescent Perceptions of Sibling Trust (APST)	191 adolescents	10-13 years old	Ethnically diverse (no exact numbers or definition) adolescents from lower to middle class schools	School-based individual administration via tablets
Jiang, Cao, Huang, Wu & Chen	2022	Frontiers in Psychology	China	Instrument Development	Sibling Relationship Questionnaire for Chinese Preschool Children (Parental Version)	591 mothers of preschool children	3-6 years old	Preschool children in China from five kindergartens	Home-based parental questionnaire with data collected at kindergartens
López-Fernández, Gomez-Benito, Kramer & Barrios	2022	Family Relations	Spain	Instrument Validation	Parental Expectations and Perceptions of Children's Sibling Relationships Questionnaire (PEPC-SRQ)	244 mothers	Mothers of 2–10-year-old children	79% of Spanish mothers were college educated, with 90% from intact families	Interviews and online questionnaires
Agnew, Fiani & Jones	2023	Behaviour Modification	United States	Instrument Development	Sibling Relationship Assessment (SRA)	40 children	4-14 years old	Sibling dyads with one sibling with ASD, no information on ethnicity or class	Research-based assessments in controlled conditions
Xu, Wang, Gao, Wang & Wu	2023	Frontiers in Psychology	China	Instrument Validation	Chinese-Sibling Inventory of Behaviour (C-SIB)	590 parents of preschool children	3-6 years old	Preschool children in China	Home-based parental questionnaire

Measure Characteristics

Table 1.2 provides an overview of synthesised key information about the measure characteristics.

Type of Report

Twelve studies included measures that were child-reported, SRQ and its validation studies, SRI and its validation studies, SAI, and SRA, and five studies included measures that were parent-reported, MIACSR, PEPC-SRQ and its validation studies, SRQ-CPC, EC-SRQ and C-SIB.

Item Count and Format

The number of items ranges from as few as 9 (APST) to 48 (SRQ). Sixteen responses were indexed using Likert–scales and one study was an observational measure that recorded the percentage of time the siblings spent in the social interaction condition.

Recurring Factors

Several factors were consistently reported across measures. Rivalry (including Rivalry/Competition and Rivalry/Jealousy) appeared in eleven studies, while Warmth (including Warmth/Closeness and Sibling Warmth) was noted in nine. Conflict (or Conflict/Rivalry) and Relative Power/Status appeared in six studies, and Affection in four. Additionally, Competition and Hostility were each reported in three studies, with Agonism, Aggression, and Jealousy noted in two studies. A range of other factors, such as Cooperation, Help, Smiles, Protests, Teasing, Admiration, Respect, Communication, Trust, Alienation, Reliability, Honesty, Companionship, Empathy, Teaching, Avoidance, Sibling Interaction, Sibling Acceptance, Social Interaction Preference, and Joint Play, were each reported in one study.

Population Focus

Fifteen studies focused on typically developing siblings, whilst two studies focused on specialised populations, such as siblings of children with traumatic brain injuries (Swift et al., 2003) or autism spectrum disorder (ASD) (Agnew et al., 2023).

Population and Translations

Instruments initially developed in North American contexts, such as the SRQ, the SRI, and the PEPC-SRQ, have been translated into Greek (Adamis et al., 2017), Dutch (Boer et al., 1997; Derkman et al., 2010), Italian (Lecce et al., 2010), and Spanish (Lopez-Fernandez et al., 2022).

Context and Application

Most measures were developed within research settings, with potential applications in clinical and educational settings.

Table 1.2

Measure Characteristics

Name	Reference	Purpose	Child or Parent Reported	Underlying Factors	Number of Items	Scoring Method	Response Options	Intended Context of Use
Sibling Relationship Questionnaire (SRQ)	Furman & Buhrmester (1985)	To measure children's perceptions of sibling relationship qualities	Child reported	Warmth/Closeness, Relative Power/Status, Conflict, Rivalry	48	Likert Scale	1 = Hardly at all, 5 = Extremely much	Within a research setting, potential clinical and educational applications
Maternal Interview About Children's Sibling Relationships (MIACSR)	Stocker, Dunn & Plomin (1989)	To create a comprehensive, multi-method assessment of sibling relationships	Parent reported	Positive (joint play, affection, cooperation/help, smiles) Negative (jealousy, aggression, competition, protests, teasing)	17	Likert Scale	1 = Almost never/rarely, 5 = Regularly/just about every day	Within a developmental research setting, conducted in the family home
Sibling Relationship Questionnaire (SRQ)	Buhrmester & Furman (1990)	To measure children's perceptions of sibling relationships during middle childhood and adolescence	Child reported	Relative Power/Status, Warmth/Closeness, Conflict/Rivalry	48	Likert Scale	1 = Hardly at all, 5 = Extremely much	Within a research setting, potential clinical and educational applications

Sibling Relationship Inventory (SRI)	Stocker & McHale (1992)	To assess children's self-reported behaviours and perceptions toward their siblings	Child reported	Affection, Rivalry, Hostility	17	Likert Scale	1 = Not at all, 5 = All the time	Within a research setting, potential clinical and educational applications
Parental Expectations and Perceptions of Children's Sibling Relationships Questionnaire (PEPC-SRQ)	Karmer & Baron (1995)	To assess parental perceptions, standards, and expectations for their children's sibling relationship behaviours	Parent reported	Warmth, Agonism, Rivalry/Competition	24	Likert Scale	1= Never, 5 = Always	Within a research and applied intervention development setting. Potential clinical and educational applications
Sibling Relationship Inventory (SRI)	Boer, Westenberg, McHale & Updegraff (1997)	Assess a shortened version of SRI in American and Dutch samples	Child reported	Affection, Hostility, Rivalry	14	Likert Scale	1 = Almost never, 5 = Always	Within a research setting, potential clinical and educational applications
Sibling Relationship Questionnaire (SRQ)	Swift, Taylor, Kaugars, Drotar, Yeates, Wade & Stancin (2003)	To assess sibling relationship quality in families of children with traumatic brain injury (TBI)	Child reported	Warmth/Closeness, Conflict, Rivalry, Admire/Respect	24	Likert Scale	1 = Hardly at all, 5 = Extremely much	Within a research setting, potential clinical and educational applications

Sibling Relationship Inventory (SRI)	Lecce, de Bernart, Vezzani, Pinto & Primi (2011)	To examine the psychometric properties of the Sibling Relationship Inventory (SRI) in an Italian sample	Child reported	Affection, Hostility, Rivalry	17	Likert Scale	1 = Not at all, 5 = All the time	Within a research setting, potential clinical and educational applications
Sibling Relationship Questionnaire (SRQ)	Derkman, Scholte, Van der Veld & Engles (2010)	To confirm the factorial structure of SRQ for internalising externalising behaviours in a sample of Dutch adolescents.	Child reported	Warmth/Closeness, Conflict	21	Likert Scale	1 = Hardly at all, 5 = Extremely much	Within a research setting, potential clinical and educational applications
Sibling Relationship Questionnaire (SRQ)	Adamis, Tsamparl & Talanti (2017)	Translate SRQ into Greek and examine its psychometric properties.	Child reported	Warmth/Closeness, Relative Status/Power, Conflict, Rivalry	21	Likert Scale	1 = Hardly at all, 5 = Extremely much	Within a research setting, potential clinical and educational applications
Sibling Attachment Inventory (SAI)	Noel, Francies & Tilley (2018)	To measure attachment components in sibling relationships and develop a measure.	Child reported	Communication, Trust, Alienation	21	Likert Scale	1 = Never true, 3 = Always true	Within a research setting, potential clinical and educational applications
Early Childhood Sibling Relationship	Wang, Li, Liu, Zhao, Li & Niu (2021)	To develop and validate a sibling relationship quality	Parent reported	Warmth, Rivalry/Jealousy, Conflict	22	Likert Scale	1 = Never happens, 5 = Always happens	Within a research setting, potential clinical and educational applications

Questionnaire (China-specific)		questionnaire in early childhood within contemporary Chinese culture.						
Adolescent Perceptions of Sibling Trust (APST)	Persram, Howe & Bukowski (2022)	To measure adolescent's perceptions of trust in their sibling relationships and develop a measure.	Child reported	Reliability, Honesty	9	Likert Scale	1 = Not at all true, 5 = Very true	Within a research setting, potential clinical and educational applications
Sibling Relationship Questionnaire for Chinese Preschool Children (Parental Version)	Jiang, Cao, Huang, Wu & Chen (2022)	To measure the structure and quality of sibling relationships in preschool-aged children in China.	Parent reported	Sibling Interaction, Sibling Acceptance, Sibling Warmth, Sibling Rivalry	18	Likert Scale	1 = Never happens, 5 = Always happens	Within a research setting, potential clinical and educational applications
FFParental Expectations and Perceptions of Children's Sibling Relationships Questionnaire (PEPC-SRQ)	Lopez-Fernandez, Gomez-Benito, Kramer & Barrios (2022)	Translate PEPC-SRQ into Spanish and examine its psychometric properties.	Child reported	Warmth, Agonism, Rivalry/Competition	24	Likert Scale	1= Never, 5 = Always	Within a research setting, potential clinical and educational applications

Sibling Relationship Assessment (SRA)	Agnew, Fiani & Jones (2023)	To develop and validate a sibling relationship assessment measure in children with ASD,	Child reported	Social Interaction Preference	NA - behavioural assessment	% of time the sibling spends in the social interaction condition	The physical location: standing, sitting/lying down	Within a research setting, potential clinical and educational applications
Chinese-Sibling Inventory of Behaviour (C-SIB)	Xu, Wang, Gao, Wang & Wu (2023)	To validate the C-SIB by examining its psychometric properties in China	Parent reported	Companionship, Empathy, Teaching, Rivalry, Aggression, Avoidance	32	Likert Scale	1 = Never, 5 = Always	Within a research setting, potential clinical and educational applications

Measurement Properties: Methodological Quality and Quality Criteria

The following sections outline the COSMIN-based assessment (Mokkink et al., 2024) of each measure's psychometric properties. See Table 1.3 for a summary of these.

PROM Development

Eight studies have reported on the development of new measures (SRQ, SRI, PEPC-SRQ, SAI, APST, PEP-SRQ-PV SRA and EC-SRQ). Of these, EC-SRQ received the highest rating *Very Good*. PEP-SRQ-PV was rated *Adequate*, SRQ was rated *Doubtful*, and SRI, PEPC-SRQ, SAI, and APST were *Inadequate* due to limited or absent qualitative methods in item generation.

Content Validity

All measures clearly described the underlying constructs and linked item content to well-established theoretical frameworks in sibling relationship literature, specifying age range, family composition, and settings in which the measures were intended. Nevertheless, SRQ, SRI, PEP-SRQ, SAI and APST received an *Inadequate* rating due to limited and/or absent qualitative methods in item development and refinement. PEP-SRQ-PV and EC-SRQ were rated *Adequate* as they partly addressed wording or general clarity from participant feedback; however, they did not probe systematically for missing concepts or participant understanding. The SRA employed direct behavioural observation instead of the qualitative item generation method, rendering standard content-validity assessment via interviews or focus groups not applicable. Few studies (SRQ; Buhrmester & Furman, 1990; SAI; Noel et al., 2018; APST; Persram et al., 2022 and SRA; Agnew et al., 2023) reported a dedicated "item relevance" step or formal cognitive interviews specifically addressing comprehensibility, describing this step minimally or not at all. Although expert input from teachers, professors, and postgraduates was sometimes sought, the process was not always highly structured or described in detail. The highest risk of bias was *Adequate* for PEP-SRQ-PV and EC-SRQ, with the remainder being *Inadequate*.

Structural Validity

Six studies, two of SRQ's validation studies (Derkman et al., 2010, Adamis et al., 2017), SRI validation study (Lecce et al., 2010), PEP-PSQ-PV, EC-SRQ and C-SIB, received a *Very Good* rating as they performed both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Four studies used only EFA, receiving *Adequate* rating (PEPC-SRQ; Kramer & Baron, 1995, Lopez-Fernandez et al., 2022; SAI; Noel et al., 2018 and APST; Persram et al., 2022), while four studies, the original SRQ (Furman & Buhrmester, 1985) and it's validation study (Buhrmester & Furman, 1990), MIACSR, original SRI (Stocker & McHale, 1992) received *Doubtful* rating as they relied solely on principal component analysis (PCA) to reduce the variables into a smaller set, rather than using exploratory or confirmatory factor analysis to examine their factor structure. Finally, the SRQ validation study (Swift et al., 2003) received an *Inadequate* rating due to performing PCA with a sample size that was less than five times the number of items in the tested model, falling below the COSMIN's recommended threshold for adequate factor analysis. Most studies met the COSMIN standards for sample size recommendation, generally achieving five to seven participants per item or more.

Internal Consistency

Internal consistency was consistently strong across measures. Fifteen studies were rated *Very Good*, MIACSR rated *Inadequate* due to neither Cronbach's alpha nor item-total correlations calculated, and SRA rated *Not Applicable* due to being a single, direct observational measure. Most used Cronbach's alpha (≥ .80 in many subscales), and one study (C-SIB; Xu et al., 2023) calculated Omega. Subscales such as Warmth/Closeness and Conflict had high alpha values (range: .84–.92), while Rivalry or Status/Power were sometimes lower but still moderate (e.g., .58–.66).

Cross-Cultural Validity

Most measures received *Not Applicable* ratings in this domain as they were validated in a single language or single cultural context without formal multi-group CFA or invariance testing. Although the SRQ, SRI and PEPC-SRQ were translated into Greek, Dutch, Italian, and Spanish, no multigroup CFA was reported, so true cross-cultural

equivalence is unclear. Boer et al. (1997) used United States and Dutch samples for their SRI validation study but did not conduct formal invariance testing, so it was rated *Inadequate*. The Chinese measures PEP-SRQ-PV, EC-SRQ and C-SIB were rated *Doubtful* as within-country invariance analysis in China demonstrated partial or full invariance across different regions, suggesting some cross-regional validity. However, no direct cross-linguistic equivalence was tested versus an English version, so international cross-cultural validity is still undetermined.

Reliability

Test-retest reliability was tested in several studies. The SRA received *a Very Good* rating, five studies, SRQ validation study (Adamis et al., 2017), SRI validation study (Boer et al., 1997), PEPC-SRQ and its validation study, and C-SIB five were *Adequate*, and five studies, SRQ, MIACSR, SRI, PEP-SRQ-PV and EC-SRQ, were *Doubtful*. Many relied on Pearson correlations over the recommended intra-class correlation coefficient (ICC). Time intervals varied (two weeks to several months), with shorter intervals generally yielding stronger correlations.

Measurement Error

Measurement error is the difference between a measured quantity and its true value. Five studies, PEPC-SRQ and its validation study, EC-SRQ, SRA and C-SIB, received an *Adequate* rating for measurement error as stability was assumed but not confirmed. Although the time interval and measurement conditions were appropriate, the MIACSR received an *Inadequate* rating as they did not report any explicit indices of measurement error. The SRI was rated as *Doubtful* due to test-retesting using Pearson's *r* instead of ICC. The remaining studies were rated *Not Applicable*.

Criterion Validity

Criterion validity is the extent to which a measurement instrument correlates with or predicts an external criterion (Mokkink et al., 2024). All studies received a *Not Applicable* rating due to the absence of a current 'gold-standard' sibling relationship measure.

Hypothesis Testing for Construct Validity

Hypothesis testing can occur by comparison to other outcome measures (convergent validity) or differences in scores between known groups (discriminative validity) (Mokkink et al., 2024). For the former, ten studies were rated *Very Good*, two were rated *Adequate,* and three were rated *Doubtful*. For the latter, four studies were rated as *Very Good*, one *Adequate,* and one *Doubtfu*l. Studies generally compared the new or adapted instrument with existing, theoretically related measures, such as SRQ questionnaires, attachment scales or behavioural checklists, with these comparator instruments being generally well described.

Responsiveness

Few measures explicitly examined responsiveness to change, for example, pre-post intervention or longitudinal studies. As a result, all instruments were rated *Not Applicable*, reflecting limited evidence of sensitivity to detect within-person changes over time.

Administration and Contextual Considerations

Certain items in parent-reported instruments may be less relevant for older children or adolescents, particularly when they focus on preschool behaviours. Most tools are straightforward to administer in a typical research or educational setting. However, their adaptability and relevance should be carefully considered when applied to specialised populations or unique family structures. The SRA for ASD populations involved direct observation rather than self-report, which required trained personnel and a specialised setup.

Table 1.3

Quality of Studies on Measurement Properties using COSMIN Risk of Bias

	Instrument	PROM Development	Content Validity	Structural Validity	Internal Consistency	Cross-Cultural Validity	Reliability	Measurement Error	Criterion Validity	Hypothesis Testing (Convergent Validity)	Hypothesis Testing (Discriminant Validity)	Responsiveness
Original Measure	**SRQ (Furman & Buhrmester, 1985)**	D	I	D	VG	NA	D	NA	NA	NA	VG	NA
Validation Studies	**SRQ (Buhrmester & Fuman, 1990)**	NA	NA	D	VG	NA	NA	NA	NA	NA	VG	NA
	SRQ (Swift et al., 2003)	NA	NA	I	VG	NA	NA	NA	NA	VG	VG	NA
	SRQ (Derkman et al., 2010)	NA	NA	VG	VG	NA	NA	NA	NA	VG	NA	NA
	SRQ (Adamis et al., 2017)	NA	NA	VG	VG	NA	A	NA	NA	A	NA	NA
Validation Study	**MIACSR (Stocker et al., 1989)**	NA	NA	D	I	NA	D	NA	NA	VG	VG	NA
Original Measure	**SRI (Stocker & McHale, 1992)**	IN	IN	D	VG	NA	D	D	NA	D	D	NA
Validation Studies	**SRI (Boer et al., 1997)**	NA	NA	D	VG	I	A	NA	NA	D	A	NA
	SRI (Lecce et al., 2010)	NA	NA	VG	VG	NA	NA	NA	NA	A	NA	NA

Original Measure	**PEPC-SRQ (Kramer & Baron, 1995)**	I	I	A	VG	NA	A	A	NA	D	NA	NA
Validation Study	**PEPC-SRQ (Lopez-Fernandez et al., 2022)**	NA	NA	A	VG	NA	A	A	NA	VG	NA	NA
Original Measures	**SAI (Noel et al., 2018)**	I	I	A	VG	NA	NA	NA	NA	VG	NA	NA
	APST (Persram et al., 2021)	I	I	A	VG	NA	NA	NA	NA	VG	NA	NA
	PEP-SRQ-PV (Jiang et al., 2022)	A	A	VG	VG	D	D	A	NA	VG	NA	NA
	EC-SRQ (Wang et al., 2022)	VG	A	VG	VG	D	D	A	NA	VG	NA	NA
	SRA (Agnew et al., 2023)	NA	NA	NA *(Single obs)*	NA *(Single obs)*	NA	VG	A	NA	VG	NA	NA
Validation Study	**C-SIB (Xu et al., 2023)**	NA	NA	VG	VG	D	A	NA	NA	VG	NA	NA

Notes:
VG = Very good, A = Adequate, D = Doubtful, I = Inadequate, NA = Not Assessed or Not Applicable (i.e., no evidence was found in the included articles).
These refer to the overall rating for that domain, as extracted from the COSMIN risk-of-bias checklist.
"SRA (Agnew et al., 2023)" is a single-behaviour observational measure: structural validity and internal consistency do not meaningfully apply, so these were marked "NA"

Discussion

This systematic review aimed to identify, describe, and evaluate measures of sibling relationship quality in children and adolescents according to the COSMIN Risk of Bias criteria. Our primary aims were to identify existing measures, describe their development and psychometric properties and provide recommendations for instrument selection and future sibling relationship measurement development. Overall, we found that no single instrument currently fulfils all COSMIN criteria to a uniformly high standard. Although many measures demonstrated moderate to good psychometric properties, especially for structural validity and internal consistency, each exhibited some limitations in other domains, such as content validity, cross-cultural validity, or responsiveness.

PROM Development and Content Validity

Content validity is considered to be the most important measurement property of a PROM as it ensures that items used in an assessment are appropriate, relevant, and representative of the construct being measures, and a lack of this affects the quality of other measurement properties (Mokkink et al., 2009; 2024). Although many sibling relationship measures utilised expert opinion for item development, overall, there was a reliance on these and literature reviews without asking the target population, such as children or parents, about item relevance, comprehensiveness and comprehensibility. For example, the SRI, PEPC-SRQ, SAI, APST, and SRA, utilised literature reviews or expert opinions without conducting systematic qualitative research such as focus groups, cognitive interviewing, or concept elicitation with the target population of siblings or parents of children. In addition, although the widely used SRQ derived their items from open ended interviews, they did not re-interview the children explicitly about item relevance. Additionally, the SRI did not report whether parents were specifically asked about instructions, clarity, or response options. The PEPC-SRQ, SAI and APST relied solely on literature reviews or face validity checks by colleagues, rather than formal cognitive interviewing or pilot work.

Structural Validity

Structural validity is the degree to which the scores of the PROM are reflective of the dimensionality of the construct being measured (DeVet et al., 2011). In contrast to content validity, structural validity emerged as a relative strength for many instruments. Multiple studies met or exceeded COSMIN recommendations by performing both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). For instance, SRQ, SRI, PEP-SRQ-PV, EC-SRQ, and C-SIB each demonstrated well-supported factor structures, lending confidence to interpretations of their subscale scores. Nonetheless, some early studies, including parts of the SRQ's validation history, and a few recent adaptations relied solely on PCA or performed EFA without subsequent CFA. This provides the risk that measurement error was not fully disentangled from true factor variance. The SRQ validation study (Swift et al., 2003) had insufficient sample sizes for PCA, fewer than five participants per item, thereby compromising the robustness of the factor solutions (Mokkink et al., 2024).

Internal Consistency and Reliability

Nearly all measures demonstrated good internal consistency, with Cronbach's alpha coefficients often exceeding .80, particularly for constructs such as Warmth/Closeness. However, dimensions like Rivalry or Status/Power occasionally yielded more modest alphas (.58–.66), suggesting that these subscales may not consistently capture the intended constructs. Test-retest reliability showed greater variability in methodological rigour. While the SRA achieved a high rating, several studies relied on Pearson correlations rather than intra-class correlation coefficients (ICCs) and did not consistently report measurement error indices (e.g., Standard Error of Measurement, Smallest Detectable Change). This limits our confidence in the stability of scores over time, particularly in contexts, for example, interventions, where detecting change is paramount. Additionally, inconsistencies in reliability testing hinder the ability to determine whether observed changes in sibling relationship quality reflect true changes or are simply due to measurement error. This is especially problematic for longitudinal research and clinical interventions, where accurate tracking of change over time is essential.

Cross-Cultural Validity

While the SRQ, SRI, PEPC-SRQ, and C-SIB have been translated into languages such as Greek, Dutch, Italian, Spanish, and Chinese, enabling their use in diverse populations, the lack of formal cross-cultural validity assessment limits confidence in their accuracy for evaluating sibling relationship quality across cultural contexts. Without rigorous testing, such as multi-group CFA, it remains uncertain whether these measures consistently capture the intended constructs across different cultural or socioeconomic groups. Formal measurement invariance testing was limited, and as a result, the stability of factor structures and the equivalence of item meanings across diverse populations remain unclear, posing challenges for their use in research and clinical trials that require culturally sensitive tools.

Responsiveness

Responsiveness, or the ability to detect change over time, was minimally examined. Few studies used longitudinal or intervention-based designs to establish whether these measures can capture shifts in sibling relationship quality. This limitation affects the utility of these measures in both clinical and research settings, particularly when evaluating the effectiveness of interventions or tracking developmental trajectories. Without evidence of responsiveness, it remains uncertain whether these tools can detect meaningful changes over time, potentially limiting their role in outcome evaluations or longitudinal studies.

Recommendations

Based on our findings, we offer the following recommendations for selecting sibling relationship measures. Clinicians should prioritise tools with strong content validity and responsiveness to monitor individual changes, particularly in therapeutic or intervention settings. Researchers should focus on measures with robust structural validity and internal consistency for reliable data in group studies, while intervention research requires instruments with proven responsiveness to detect pre- and post-intervention changes. Measures should use clear, age-relevant language and offer parent-report versions or simplified formats for younger children. For content validity, the EC-SRQ and PEP-SRQ-PV show stronger qualitative grounding but still require more

systematic probing of item relevance and clarity. For structural validity, instruments that applied both EFA and CFA, such as SRQ, SRI, PEP-SRQ-PV, EC-SRQ, and C-SIB offer more robust factor structures. Researchers seeking parent-reports or shorter subscales, for example trust or attachment, may find other instruments more suitable. Ultimately, the chosen measure should align with the age group, cultural context, and key constructs of interest, for example, trust, conflict, warmth, while considering each measure's psychometric strengths and limitations.

Strengths and Limitations

A strength of this review is its comprehensive search strategy that spanned six databases, allowing for a broad capture of both widely used and emerging sibling measures. Using the COSMIN checklist provided a transparent framework to evaluate each instrument's methodological quality. Nevertheless, limitations include potential publication bias due to the exclusion of grey literature. Some studies lacked sufficient methodological detail, item development processes, rendering precise COSMIN ratings challenging. Moreover, restricting the scope to child- and adolescent-focused measures may limit applicability to adult sibling relationships or more complex family structures such as children in blended families or foster care.

Future Research Directions

Looking ahead, there is a clear need to develop or validate sibling relationship measures that follow rigorous, modern psychometric guidelines. Potential avenues include the development and validation of a new measure that includes children or parents of children across multiple developmental stages. Conducting comprehensive PROM development to ensure content validity and capture emerging relational constructs in sibling relationships, particularly in diverse or non-traditional family structures. Additionally, multi-group CFA or invariance testing could be used to confirm cross-cultural stability of existing measures. Finally, and testing responsiveness in longitudinal, clinical, or intervention studies would help determine whether sibling relationship measures can detect meaningful changes over time.

References

Adamis, D., Tsamparl, A., & Talanti, K. (2017). Psychometric analysis of the Greek version of the Sibling Relationship Questionnaire. *Psychology: The Journal of the Hellenic Psychological Society, 22(1),* 1–13. https://doi.org/10.12681/psy_hps.23175

Agnew, C. N., Fiani, T., & Jones, E. A. (2023). What if you can't ask them? Psychometric properties of a behavioral assessment of sibling relationship quality. *Behavior Modification, 47*(3), 670–692. https://doi.org/10.1177/01454455221133003

Boer, F., Westenberg, P. M., McHale, S. M., Updegraff, K. A., & Stocker, C. M. (1997). The factorial structure of the Sibling Relationship Inventory (SRI) in American and Dutch samples. *Journal of Social and Personal Relationships, 14*(6), 851–859. https://doi.org/10.1177/0265407597146004

Buhrmester, D., & Furman, W. (1990). Perceptions of sibling relationships during middle childhood and adolescence. *Child Development, 61*(5), 1387–1398. https://doi.org/10.2307/1130750

Buist, K. L., & Vermande, M. (2014). Sibling relationship patterns and their associations with child competence and problem behavior. *Journal of Family Psychology*, 28, 529–537. https://doi.org/10.1037/a0036990

De Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine*. Cambridge, United Kingdom: Cambridge University Press.

Dunn, J. (2006). Siblings and socialization. In J. E. Grusec & P. D. Hastings (Eds.), Handbook of socialization: Theory and research (pp. 309–327). New York: Guilford.

Furman, W., & Buhrmester, D. (1985). Children's Perceptions of the Qualities of Sibling Relationships. *Child Development*, 56(2), 448–461. https://doi.org/10.2307/1129733

Hetherington, E. M. (1989). Coping with family transitions: Winners, losers, and survivors. *Child Development, 60*(1), 1–14. https://doi.org/10.2307/1131066

Holmes, M. R., Bender, A. E., O'Donnell, K. A., Miller, E. K., & Conard, I. T. (2024). Illuminating the landscape of sibling relationship quality: An evidence and gap map. *Child Development*. https://doi.org/10.1111/cdev.14065

Jiang, M., Cao, X., Huang, Q., Wu, S., & Chen, X. (2022). Exploring the structure of sibling relationships among preschool children in China and developing a questionnaire. *Frontiers in Psychology, 12*, Article 745165. https://doi.org/10.3389/fpsyg.2021.745165

Kramer, L., & Baron, L. A. (1995). Parental perceptions of children's sibling relationships. *Family Relations: An Interdisciplinary Journal of Applied Family Studies, 44*(1), 95–103. https://doi.org/10.2307/584746

Kramer, L., & Conger, K. J. (2009). What we learn from our sisters and brothers: For better or for worse. *New Directions for Child and Adolescent Development, 126*, 1–12. https://doi.org/10.1002/cd.253

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

Lecce, S., de Bernart, D., Vezzani, C., Pinto, G., & Primi, C. (2011). Measuring the quality of the sibling relationship during middle childhood: The psychometric properties of the Sibling Relationship Inventory. *European Journal of Developmental Psychology, 8*(4), 423–436. https://doi.org/10.1080/17405629.2010.530033

López-Fernández, G., Gómez-Benito, J., Kramer, L., & Barrios, M. (2022). Spanish validation of the Parental Expectations and Perceptions of Children's Sibling Relationships Questionnaire. *Family Relations.*https://doi.org/10.1111/fare.12699

McHale, S. M., Updegraff, K. A., & Whiteman, S. D. (2012). Sibling relationships and influences in childhood and adolescence. *Journal of Marriage and Family*, 74(5), 913–930. https://doi.org/10.1111/j.1741-3737.2012.01011.x

Milesky, A., & Levitt, M. (2005). Sibling support in early adolescence: Buffering and compensation across relationships. *The European Journal of Developmental Psychology, 2(3),* 299–320. https://doi.org/10.1080/17405620544000048

Mokkink, L. B., Elsman, E. B. M., & Terwee, C. B. (2024). The COSMIN guideline for systematic reviews of patient-reported outcome measures (PROMs). *Quality of Life Research, 33*(10), 2929–2939. https://doi.org/10.1007/s11136-024-03761-6

Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, *18*(3), 313–333. https://doi.org/10.1007/s11136-009-9451-9

Noel, V. A., Francis, S. E., & Tilley, M. A. (2018). An adapted measure of sibling attachment: Factor structure and internal consistency of the Sibling Attachment Inventory in youth. *Child Psychiatry & Human Development, 49*(2), 217–224. https://doi.org/10.1007/s10578-017-0742-z

Persram, R. J., Howe, N., & Bukowski, W. M. (2022). Constructing and validating a measure of adolescents' perceptions of sibling trust. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 54*(1), 73–78. https://doi.org/10.1037/cbs0000281

Schaefer, E. S., Edgerton, M. (*1981*). *The Sibling Inventory of Behavior.* Chapel Hill: University of North Carolina.

Scholte, R. H. J., Engels, R. C. M. E., & Van der Veld, W. M. (2010). Factorial and construct validity of the Sibling Relationship Questionnaire. *European Journal of Psychological Assessment, 26*(1), 55–62. https://doi.org/10.1027/1015-5759/a000037

Stocker, C. M., & McHale, S. M. (1992). The nature and family correlates of preadolescents' perceptions of their sibling relationships. *Journal of Social and Personal Relationships, 9*(2), 179–195. https://doi.org/10.1177/0265407592092003

Swift, E. E., Taylor, H. G., Kaugars, A. S., Drotar, D., Yeates, K. O., Wade, S. L., & Stancin, T. (2003). Sibling relationships and behavior after pediatric traumatic brain injury. *Journal of Developmental and Behavioral Pediatrics, 24*(1), 24–31. https://doi.org/10.1097/00004703-200302000-00004

Tucker, C. J., & Finkelhor, D. (2017). The state of interventions for sibling conflict and aggression: A systematic review. *Trauma, Violence & Abuse*, 18, 396–406. https://doi.org/10.1177/15248380156224

Volling, B. L. (2003). Sibling relationships. In M. H. Bornstein, L. Davidson, C. L. M. Keyes, & K. A. Moore (Eds.), *Well-being: Positive development across the life course* (pp. 205–220). Lawrence Erlbaum Associates Publishers.

Wang, Y. Z., Li, Y., Liu, T., Zhao, J., & Li, Y. (2021). Understanding relationships within cultural contexts: Developing an early childhood sibling relationship questionnaire in China. *Family Relations.* https://doi.org/10.1111/fare.12586

Xu, H., Wang, Z., Gao, X., Wang, X., & Wu, Q. (2023). Psychometric validation of the sibling inventory of behavior in three- to six-year-old Chinese children. *Frontiers in Psychology, 14,* 1124518. https://doi.org/10.3389/fpsyg.2023.1124518

AI Software

Grammarly. (2025). *Grammarly AI* (Version 2.0). Grammarly, Inc. https://www.grammarly.com

The Grammarly Premium AI tool was used to check grammar and style to help improve the overall flow of writing for this piece of work. This included features such as review checking – which provides feedback and suggestions on correctness, clarity, engagement and writing delivery.

Chapter 2

Development and Validation of a New Measure of Sibling Relationship Quality

Prepared in accordance with the author requirements for Family Relations Journal; https://onlinelibrary.wiley.com/page/journal/17413729/homepage/forauthors.html

Plain Language Summary

Title

Development and Validation of a New Measure of Sibling Relationship Quality.

Background

The sibling relationship is important but under-studied. Current questionnaires may miss key parts of this relationship or not follow modern guidelines for creating good-quality sibling relationship measures.

Aims

This study created and tested a new questionnaire called the Sibling Relationship Quality-Informant Report (SRQ-IR). It aimed to improve existing questionnaires by looking at more sibling interactions and including children of different ages and family situations.

Methods

Researchers reviewed existing questionnaires and studies, and worked with parents, clinicians, and experts to design a compressive set of questions about sibling relationships. Then, 457 parents filled out our online questionnaire about their children's sibling dynamics over the past month. The data was analysed to find common patterns in these sibling relationships to create the final questionnaire. We also checked whether our newly developed questionnaire captured what it was meant to be measuring and whether our questions were working well together to measure the same ideas.

Results

The study found six key parts of sibling relationships: Antagonism (winding each other up), Warmth/Closeness (enjoying time together and feeling close) Relational Aggression (hurting feelings through behaviours), Conflict Resolution (solving disagreements), Mentalising (understanding each other's thoughts and feelings), and Rivalry (competing or comparing to each other). The final questionnaire found that

sibling relationships are both caring and challenging. Our checks found that our new questionnaire, the SRQ-IR, worked well for measuring sibling relationships from an adult's perspective and that our questions worked well together. We also captured new ideas about sibling relationships, like how siblings solve disagreements and how well they understand each other's thoughts and feelings.

Discussion

The SRQ-IR is a promising new sibling relationship questionnaire that can be completed by adults familiar with a sibling pair that looks at sibling relationships in many ways. It captured both positive and more challenging parts of the sibling relationship. Future studies could test this new questionnaire with more people from different backgrounds to make sure it works well for everyone, including the opinions of children and whether this new questionnaire can track changes in sibling relationships over time. This could help researchers, clinicians, and families better understand and support healthy sibling relationships.

Abstract

The Sibling relationship is important, yet it remains under-studied. Existing measures are limited to specific developmental stages, may overlook more complex relational dynamics in addition to warmth, conflict and rivalry and do not meet modern psychometric recommendations. This study aimed to develop and validate a new measure of sibling relationship quality to address these limitations. An item pool was generated and refined from existing literature and measures and through expert consultation with parents, clinicians, and researchers. The final Sibling Relationship Quality-Informant Report (SRQ-IR) was completed by 457 parents of children under 18, and exploratory factor analysis with parallel analysis was used to determine the best-fitting latent factor structure. The results yielded a robust six-factor solution comprising *Antagonism, Warmth/Closeness, Relational Aggression, Conflict Resolution, Mentalising,* and *Rivalry*, accounting for 67.74% of the variance. Reliability indices were strong (Cronbach's alphas ranged from .82 to .93), supporting the instrument's internal consistency and construct validity. The SRQ-IR provides a promising, multidimensional framework for assessing sibling relationship quality. Future research should confirm the SRQ-IR's factor structure using confirmatory factor analysis, incorporate multi-informant designs, and validate the measure across diverse samples to establish its robustness.

Introduction

Sibling relationships are among the most enduring and influential in an individual's life, shaping social, emotional, and cognitive development. These relationships serve as a training ground for interpersonal skills, fostering security and resilience in the face of challenges such as poverty, marital conflict, and family instability (Hetherington, 1989; Milesky & Levitt, 2005). Sibling relationships that are close, warm, and nurturing help children develop social competence, enhance their conflict resolution abilities, and promote their social and emotional understanding (Hughes et al., 2018; Kramer, 2014). While parent-child and peer relationships have received substantial research attention, sibling dynamics remain comparatively understudied (McHale, Updegraff, & Whiteman, 2012). Traditionally confined to biological siblings, the concept of siblinghood has evolved alongside changes in family structures, encompassing half-siblings, stepsiblings, and adoptive siblings, reflecting broader societal and cultural shifts (Ben-Arieh et al., 2014; Cicirelli, 1995).

Current Measures of Sibling Relationships

The persistent challenge in sibling research is the absence of a clearly defined and comprehensive "sibling theory" (Caspi, 2011). Although family systems theory, social learning, and attachment perspectives have guided some of the work (Whiteman, McHale & Soli, 2011), no single theoretical framework fully captures the complexity of sibling interactions. Efforts to quantify sibling relationship quality have led to the development and validation of several psychometric tools since the 1980s, consistently highlighting *Warmth*, *Conflict*, and *Rivalry* as core domains (Furman & Buhrmester, 1985, Buhrmester & Furman, 1990; Kramer & Baron, 1995; Swift et al., 2003; Derkman et al., 2010; Adamis, Tsamparl & Talanti, 2017; López-Fernández et al., 2022; Wang et al., 2021). *Warmth* typically includes affection, support, and closeness; *Conflict* typically involves aggression, hostility, and disagreements; and *Rivalry*, distinct from conflict, captures competition and perceived inequality between siblings (Buist, Deković, & Prinzie, 2013).

Instruments developed are typically child-reported (Sibling Relationship Questionnaire, SRQ; Furman & Buhrmester, 1985; Sibling Relationship Inventory, SRI; Stocker & McHale, 1992; Sibling Attachment Inventory, SAI; Noel, Francis & Tilley, 2018; Adolescent Perceptions of Sibling Trust, APST; Persram, Howe & Bukowski, 2021) or parent-reported (Parental Expectations & Perceptions of Children's Sibling Relationships Questionnaire, PEPC-SRQ; Kramer & Baron, 1995; Sibling Relationship Questionnaire for Chinese Preschool Children Parental Version, SRQ-CPC; Jiang et al., 2022; Early Childhood Sibling Relationship Questionnaire, EC-SRQ; Wang et al., 2021; Chinese Sibling Inventory of Behaviour, C-SIB; Xu et al., 2023).

The Need for A New Measure

While existing sibling relationship measures have contributed significantly to the field, these are subject to several important limitations. Firstly, most measures have been developed and validated for specific developmental stages of pre-school, middle childhood, or adolescence rather than capturing sibling relationships throughout childhood and adolescence. This an important limitation because existing individual development and validation studies suggest that the commonly recurring constructs, *Warmth*, *Conflict*, and *Rivalry,* remain stable across developmental periods. Secondly, many tools conceptualise sibling relationship quality as either positive or negative, potentially overlooking the multifaceted nature of sibling interactions, where constructs like *Conflict* and *Rivalry* may coexist with *Warmth* without necessarily indicating a negative overall dynamic. As our understanding of pro-social relationships has evolved, emerging relational constructs, such as social and emotional understanding, conflict management, and forming neutral or positive attributions (Kramer, 2010), remain underrepresented in sibling measures. An updated measure that includes these prosocial domains could aid in identifying modifiable components for targeted interventions that enhance sibling relationship quality (Holmes et al., 2024).

Third, existing instruments have been predominantly developed and validated within White, Western, middle-class, highly educated, two-parent family samples (Holmes et al., 2024), making them less applicable to diverse family structures, including

stepsiblings, half-siblings, and adoptive siblings. Limited cross-cultural validation further restricts the generalisability of sibling relationship assessments, as most measures were designed for traditional nuclear families and have not been widely tested in various cultural contexts (Ben-Arieh et al., 2014; Cicirelli, 1995). Recent research (Holmes et al., 2024) highlights the need for more inclusive, multidimensional sibling relationship quality measures that incorporate emerging or underrepresented constructs and reflect modern family structures. Finally, advancements in psychometric scale development highlight the necessity of rigorous item generation, validation, and practical usability (Morgado et al., 2017; Boateng et al., 2018). Many of the existing scales were not developed in line with best practice guidelines for scale development and validation (REFS). For example, the most widely used measure, the Sibling Relationship Questionnaire (SRQ; Furman & Buhrmester, 1985), was developed using principal component analysis, whereas contemporary standards recommend more robust exploratory factor analyses and parallel analyses to establish the optimal number of factors (Boateng et al., 2018).

Aims

This study aims to develop and validate a new, comprehensive measure of sibling relationship quality in line with contemporary psychometric standards. While enhancing cultural and structural inclusivity is not the primary focus, this study seeks to improve cross-cultural validity by recruiting a globally diverse participant sample via online data collection, extending beyond the Western, highly educated samples commonly used in prior sibling research. Moreover, by including non-traditional family structures, such as stepsiblings, half-siblings, foster, and adoptive siblings, this study may broaden the measure's relevance across a broader range of sibling experiences than previously explored.

Objectives

1. Collate and refine items from existing literature, sibling relationship measures and through expert consultation with parents, clinicians, and researchers.
2. Conduct an exploratory factor analysis (EFA) and a parallel analysis (PA) to identify the core dimensions of sibling relationships.
3. Evaluate the psychometric properties of the new measure, including internal consistency, convergent validity, discriminant validity, and test-retest reliability.
4. Assess the concordance of the identified factor structures with previous research findings.
5. Provide recommendations for further development and validation of sibling relationship measures.

By adhering to modern psychometric standards, this study aims to produce a developmentally inclusive, culturally adaptable tool that reflects the complexity inherent in sibling relationships, thereby advancing both empirical research and practical interventions in this domain.

Methods

Participants

This study comprised three phases. Phase 1 involved research team members (ER, AS, HM) developing an initial item pool based on a literature review, existing measures of sibling relationship quality, and adapting items from other tools measuring related relational constructs such as emotional regulation, mentalisation, and conflict resolution. Phase 2 involved parents, clinicians, and researchers refining the item pool. Fifteen parents/guardians of at least two children under 18, 28 clinicians working with children and families, and 11 researchers/clinician-academics with relevant publications in the past five years provided feedback on the item pool. Phase 3 involved a large (n = 457), heterogeneous sample of parents and guardians of children under 18 who completed the final item pool. Two hundred and forty-five parents (53.6%) were female, and 10 (2.2%) did not report their gender. The mean age of the parents was 35 years (M = 35.96, SD = 6.91). Further demographic details for Phase 3 are in the Results section.

Ethics

Ethical approval was granted by the College of Medical, Veterinary, and Life Sciences Ethics Committee, University of Glasgow (ID: 200230225) in May 2024 (Appendix 2.3). Participants in Phases 2 and 3 received study information, privacy notices, and the opportunity to ask questions before providing informed consent. They were informed of their right to withdraw at any time. Data handling was compliant with GDPR (2018) rules. Personal data were stored electronically in encrypted, password-protected files on secured University of Glasgow computers, accessible only to the primary researcher (ER) and supervisors (AS, HM). Anonymised data were used for analysis to ensure confidentiality. To minimise potential risks if respondents became distressed, they were given information about sources of support, and the need to address participant wellbeing and safety issues was reviewed throughout the study.

Materials

Participant information sheets, consent forms, demographic questions, and the newly developed Sibling Relationship Quality–Informant Report (SRQ-IR) were all delivered online via the Qualtrics survey platform. The SRQ-IR employs a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) to assess sibling interactions, attitudes, and behaviours based on parental observations of two target children over the past month. Some items feature dynamic text-entry fields (e.g., “One orders the other around (child one or child two?)”) to capture nuanced sibling dynamics. Subscales identified through exploratory factor analysis yield individual scores, with higher subscale scores indicating a greater presence of each measured construct. See Appendix 2.4 for participant information sheets and consent forms and Appendix 2.5 for the SRQ-IR’s instructions, items, and scoring details.

Procedures

Participants for Phases 2 and 3 were recruited through social media platforms such as Facebook, Instagram, LinkedIn, Reddit and targeted outreach to parenting networks and sibling researchers identified through a literature review. Study advertisements provided a brief overview, eligibility criteria, and a survey link. In Phase 2, parents,

clinicians, and researchers first completed screening questions. They then reviewed an initial item pool organised by thematic domains and offered free-text feedback on item clarity, additions, or deletions. The research team systematically evaluated these suggestions to improve item relevance and comprehensiveness. In Phase 3, parents and guardians completed screening questions, followed by a demographic questionnaire assessing their age, gender, ethnicity, education, marital status, income, and the demographics of their children. Finally, eligible parents and guardians completed the final, 187-item, SRQ-IR questionnaire.

Data Analysis

Data were collected online from July to November 2024. In Phase 2, feedback was gathered from 15 parents, 28 clinicians, and 11 researchers. In Phase 3, 641 participants initially completed the SRQ-IR questionnaire; however, 119 (18.6%) were excluded due to incomplete responses, 38 for having children over 18, and 40 for careless responding (completing the questionnaire in under five minutes), leaving a final sample of 457. Data were analysed using SPSS (Version 29; IBM, 2023), with listwise deletion applied to missing data, yielding a valid N ranging from 411 to 452 across analyses. An EFA was conducted to determine the factor structure of the measure. Parallel analysis (PA; Velicer, Eaton & Fava, 2000; Zwick & Velicer, 1986) was used to inform the number of factors. Maximum Likelihood extraction enhanced statistical robustness, and Promax rotation was applied to account for potential factor correlations (Tabachnick & Fidell, 2019). These analyses provided an initial validation of the measure and informed its refinement. See Appendix 2.6 for the data analysis plan.

Results

In Phase 2, most parents (86.3%) identified their children as biological siblings. Most clinicians (71.4%) were psychologists, and a similar proportion (71.4%) worked in the public sector. Researchers held various roles, with the largest group (36.4%) being higher education lecturers, senior lecturers, professors, or readers. See Table 2.1 for a summary of these demographic characteristics.

Table 2.1

Demographic Characteristics of Participants in Phase 2

Sample	Category	n	%
Parents: sibling relationship	Biological siblings	13	86.3
	Half siblings	2	13.3
	Total	15	
Clinician's Professions	Psychologist	20	71.4
	Nurse	2	7.1
	Social Worker	1	3.6
	Speech & Language Therapist	1	3.6
	Healthcare Support Worker	1	3.6
	Family Support Worker	1	3.6
	Missing	2	7.1
Type of Clinical Work	Public sector	20	71.4
	Private sector	7	25
	Other	1	3.6
	Total	28	
Researcher's Occupation	Research assistant	2	18.2
	PhD student or researcher	1	9.1
	Research Associate Or Fellow	2	18.2
	Higher Education Lecturer, Senior Lecturer, Professor Or Reader	4	36.4
	Missing	2	18.2
	Total	9	

Phase 3 included 245 mothers and 202 fathers aged 25 to 55 years, with 79.0% aged 25 to 45 and one missing age response. Slightly more participants identified as female (53.6%) than male (44.2%), with 2.2% missing gender data. The majority were White (79.6%), followed by Black/African/Caribbean (11.2%), with smaller proportions identifying being from Other, Mixed, or Asian backgrounds. Educational attainment was high, with 45.2% holding an undergraduate degree and 25.6% a master's degree. Most participants were married (74.8%), and 28.1% reported an annual income of £40,000–£59,999. Ninety percent of parents were biologically related to Child 1 and 87.7% to Child 2. Child 1 and Child 2 were aged between 1 to 18 years with the 50% of Child 1 being between 12 to 18 years old (M = 11.04, SD = 4.14). Child 2 were generally younger,

with 48% being between 6 to 11 years old (*M* = 7.86, *SD* = 3.86). Child 1 was mostly male (63.5%), while Child 2 had a nearly even gender split (49.7% male, 48.1% female). Most children were White (78.1% for Child 1, 74.0% for Child 2), and 93.9% lived together. See Table 2.2 for a summary.

Table 2.2

Demographic Characteristics of Participants in Phase 3

Variable	Category	n	%
Age	18–24 years	12	2.63
	25–34 years	171	37.42
	35-44 years	209	45.73
	45-55 years	64	14.00
	Missing	1	0.22
Gender (Participants)	Male	202	44.2
	Female	245	53.6
	Missing	10	2.2
Ethnicity (Participants)	White	364	79.6
	Black/African/Caribbean	51	11.2
	Asian	11	2.4
	Mixed	8	1.8
	Other	18	3.9
	Prefer to self-describe	1	0.2
	Prefer not to say	1	0.2
	Missing	3	0.7
Highest Education	Undergraduate degree	208	45.2
	Master's degree	117	25.6
	Secondary school	79	17.3
	Doctorate/PhD	32	7.0
	Primary school	3	0.7
	Missing	10	2.2
Marital Status	Married	342	74.8
	Living with partner	41	9.0
	Divorced/Separated	11	2.4

Variable	Category	n	%
	Never married	5	1.1
	Missing	58	12.7
Household Income	£40,000–£59,999	131	28.1
	More than £100,000	103	22.5
	£60,000–£99,999	99	21.7
	£20,000–£39,999	76	16.6
	Less than £20,000	30	6.8
	Missing	18	3.9
Relationship to Child 1	Biological parent	412	90.0
	Step-parent	15	3.3
	Adoptive parent	8	1.8
	Foster parent	7	1.5
	Guardian	4	0.9
	Missing	11	2.4
Age of Child 1	1–5 years	49	10.7
	6–11 years	179	39.2
	12–18 years	229	50.1
Gender of Child 1	Male	290	63.5
	Female	157	34.4
	Transgender	1	0.2
	Missing	9	2.0
Ethnicity of Child 1	White	357	78.1
	Black	50	10.9
	Mixed	9	2.0
	Other	16	3.5
	Prefer not to say	1	0.2
	Missing	14	3.1
Relationship to Child 2	Biological parent	401	87.7
	Foster parent	19	4.2
	Adoptive parent	9	2.0
	Step-parent	8	1.8
	Guardian	6	1.3
	Missing	14	3.1
Age of Child 2	1–5 years	151	33.0
	6–11 years	223	48.8
	12–18 years	83	18.2

Variable	Category	n	%
Gender of Child 2	Male	227	49.7
	Female	220	48.1
	Missing	10	2.2
Ethnicity of Child 2	White	338	74.0
	Black	50	10.9
	Mixed	19	4.2
	Other	16	3.5
	Asian	10	2.2
	Missing	24	5.3
Living Arrangements	Child 1 and 2 live together	429	93.9
	Do not live together	16	3.5
	Other	3	0.7
	Missing	9	2.0

Exploratory Factor Analysis

An EFA was conducted on the 187-item pool using Maximum Likelihood (ML) extraction with Promax rotation. Inspection of the correlation matrix revealed all items correlated r = .3 or more, meeting the minimum threshold for factor analysis. (Field, 2013). Bartlett's test confirmed that an EFA was appropriate for the sample $\chi^2(17391) = 61804.95$, $p < .001$, and a Keiser-Meyer-Olkin (KMO) test indicated an adequate participant-to-item ratio with KMO = .86. A parallel analysis was conducted to establish how many factors should be extracted using O'Connors syntax (O'Connor, 2000). Parallel analysis creates data sets with the same number of cases and variables as the actual dataset, filled with random numbers. The first six eigenvalues (and % of variance accounted for) extracted for 95% of the simulated data set, suggesting six factors are meaningful and should be retained for further analysis.

A second EFA was then conducted to force the extraction of six factors using Promax rotation, which allows factors to correlate and reflects the underlying theory that psychological constructs are expected to be interrelated. The parallel analysis results combined with EFA suggested that six factors best represent items from the Sibling

Relationship Quality-Informant Report (SRQ-IR) scale. The six-factor solution accounted for 67.74% of variance initially (30.48%, 19.84%, 5.61%, 4.55%, 4.23%, and 3.03%) and 60.85% after extraction. Following rotation, variance was more evenly distributed across factors: Factor 1 (7.58%), Factor 2 (7.63%), Factor 3 (6.17%), Factor 4 (6.21%), Factor 5 (6.13%), and Factor 6 (6.22%). Inspection if the items on each factor suggest that Factor 1 represents *Antagonism*, Factor 2 represents *Warmth/Closeness*, Factor 3 represents *Relational Aggression*, Factor 4 represents *Conflict Resolution*, Factor 5 represents *Mentalising*, and Factor 6 represents *Rivalry*.

Item reduction and subscale development

The final 34-item SRQ-IR questionnaire was derived using the following criteria, which were adopted to extract the greatest number of factors that would be well-defined and reasonably distinct from one another. When two items correlated strongly (>0.75), the item with the lowest item total correlation was considered to contain redundant information and was deleted (Clark & Watson, 1995). Items with loadings <0.40 or which demonstrated reasonably strong loadings (>0.3) on more than one factor were eliminated to maximise individual subscales' measurement properties and discriminant validity (Clark & Watson, 1995). See Tables 2.3 for a summary of SRQ-IR's factors, sample items and reliability and Table 2.4 a summary of the EFA results.

Table 2.3

Summary of SRQ-IR Factors, Sample Items, and Reliability

Factor	No. of Items	Sample Item	Cronbach's Alpha
Factor 1 - Antagonism	6	They deliberately annoy each other	0.93
Factor 2 – Warmth/Closeness	6	They choose to spend their free time together	0.82
Factor 3 – Relational Aggression	6	One attempts to sabotage the other	0.93
Factor 4 – Conflict Resolution	6	They are able to find compromises together	0.89
Factor 5 – Mentalising	6	They are interested in guessing what the other is feeling	0.87
Factor 6 – Rivalry	4	One thinks the other is more successful	0.87

Table 2.4

Summary of EFA Results for the SRQ-IR Questionnaire (N = 457)

		Rotated Factor Loadings					
Item	Mean (SD)	Antagonism	Warmth/Closeness	Relational Aggression	Conflict Resolution	Mentalising	Rivalry
1. They grass each other up (tell on each other)	2.55 (1.41)	**.83**	.13	.06	.08	-.01	-.03
2. They get on each other's nerves	2.53 (1.43)	**.81**	-.07	.04	-.01	.06	.01
3. They argue with each other	2.67 (1.39)	**.79**	-.01	-.02	-.11	.09	.09
4. They tease each other	2.60 (1.44)	**.77**	.08	.10	-.01	-.05	.02
5. They deliberately annoy each other	2.52 (1.39)	**.75**	-.08	.14	.13	-.08	-.06
6. They upset each other	2.43 (1.36)	**.68**	-.07	.18	.03	.03	-.03
7. They like each other	4.17 (1.02)	.10	**.75**	-.05	-.01	-.00	-.09
8. They choose to spend their free time together	4.09 (1.02)	-.07	**.75**	.16	-.03	-.01	-.14
9. Difficult times bring them closer together	4.08 (1.03)	-.06	**.67**	-.00	.08	-.02	.03
10. They take care of each other	4.16 (0.96)	.042	**.64**	-.10	.12	-.01	.06
11. They act as if they are on the same team	4.04 (1.05)	-.07	**.57**	-.05	-.02	.20	.08
12. They are physically affectionate with each other	3.75 (1.24)	0.026	**.53**	.06	-.09	-.02	.16
13. One attempts to turn other children against the other (child one or two?)	2.03 (1.29)	.04	-.01	**.83**	.11	-.08	.01
14. One displays developmentally inappropriate sexual behaviours towards the other (child one or two?)	1.99 (1.34)	-.06	.03	**.81**	-.08	.07	.06

15. One attempts to sabotage the other (child one or two?)	2.10 (1.28)	.08	.04	**.79**	-.03	-.04	.03
16. One can be unnecessarily cruel to the other (child one or two?)	2.10 (1.32)	.11	-.01	**.78**	-.05	.01	-.04
17. One causes fear and alarm towards the other (child one or two?)	2.08 (1.27)	.13	-.00	**.76**	-.03	.04	.00
18. One makes the other say or do things they do not want to do (child one or two?)	2.06 (1.28)	.15	.00	**.67**	-.06	0.12	.07
19. They accept apologies from each other after an argument	3.98 (1.11)	.09	-.04	-.08	**.84**	-.04	.02
20. They try to listen to each other during a conflict	3.70 (1.25)	-.30	-.02	.34	**.76**	-.01	-.04
21. They do not need a lot of prompting to apologise to each other after an argument	3.73 (1.22)	-.06	.05	.02	**.73**	.07	-.01
22. They are able to find compromises together	3.82 (1.16)	-.01	-.01	.02	**.71**	.07	-0.04
23. They are able to be friendly with each other once after an argument has been resolved	4.00 (1.10)	.22	0	-.27	**.70**	.07	.09
24. They seek help from adults or older children when conflicts become challenging to resolve	3.98 (1.16)	.11	.04	-.09	**.63**	.079	-.02
25. They try to think about how each other feels	3.77 (1.13)	-.03	.02	-.03	.01	**.81**	-.04
26. They try to think about the reasons why each other behaves the way they do	3.69 (1.16)	.06	-.02	.06	.05	**.72**	-.05
27. They know why each other acts the way they do	3.72 (1.14)	.04	0.04	.04	.05	**.69**	.02
28. They try to see situations through the eyes of the other	3.65 (1.17)	.08	-.10	.08	.03	**.66**	-.01
29. They try to understand what each other wants	3.74 (1.14)	-.02	.10	-.02	-.01	**.62**	-.08
30. They are interested in guessing what the other is feeling	3.75 (1.13)	-.03	-.08	-.03	.13	**.61**	.14
31. One thinks the other is more talented (child one or two?)	2.63 (1.33)	.05	.08	.07	.01	-.05	**.73**
32. One thinks the other is more successful (child one or two?)	2.54 (1.36)	.01	-.04	0.14	-.00	.04	**.72**

33. One thinks the other is better (child one or two?)	2.61 (1.38)	.02	-.01	.08	-.06	.09	**.72**
34. One feels overshadowed by the other (child one or two?)	2.52 (1.33)	-.02	.06	.23	.10	-.14	**.66**
Eigenvalues		10.36	6.75	1.91	1.55	1.44	1.03
% of variance		30.48	19.84	5.61	4.55	4.23	3.03
A		.93	.82	.93	.89	.87	.87

Validity

All 34 items had standardised factor loadings >0.5, indicating strong explanatory power (Hair et al., 2010). Factor loadings ranged from 0.53–0.84 across dimensions: *Antagonism* (0.68–0.83), *Warmth/Closeness* (0.53–0.75), *Relational Aggression* (0.67–0.83), *Conflict Resolution* (0.63–0.84), *Mentalising* (0.61–0.81), and *Rivalry* (0.66–0.73). Average Variance Extracted (AVE) values were 0.598 (*Antagonism*), 0.432 (*Warmth/Closeness*), 0.601 (*Relational Aggression*), 0.535 (*Conflict Resolution*), 0.474 (*Mentalising*), and 0.501 (*Rivalry*). The conventional threshold for AVE for Confirmatory Factor Analysis (CFA) AVE ≥ 0.50 indicates strong convergent validity (Fornell & Larcker, 1981), however values under ≥ 0.50 are acceptable in exploratory research as contextual and practical considerations outweigh a strict numerical cut off (Costello & Osborne, 2005).

Discriminant validity was assessed through correlations between dimensions and the total score. Inter-dimension correlations ranged from -0.027 to 0.643, all below the 0.85 threshold, indicating no significant factor overlap (Hair et al., 2010). The highest correlation (0.643) was between *Antagonism* and *Rivalry*. Additionally, the square roots of AVE values (0.657–0.775) exceeded inter-factor correlations, confirming that each dimension is distinct and measures a unique construct (Fornell & Larcker, 1981; Hair et al., 2010).

Reliability

Cronbach's alpha was calculated for Phase 3 parents (n = 457), showing strong internal consistency across SRQ-IR factors: *Antagonism* (α = .93), *Warmth/Closeness* (α = .82), *Relational Aggression* (α = .93), *Conflict Resolution* (α = .89), *Mentalising* (α = .87), and *Rivalry* (α = .87). Split-half reliability was assessed by dividing the 34-item scale into two 17-item halves. Cronbach's alpha for Part 1 (α = .843) and Part 2 (α = .840) indicated strong consistency. The correlation between halves was r = .424, with Spearman-Brown and Guttman Split-Half coefficients of 0.596 and 0.595, respectively, demonstrating moderate reliability, acceptable for exploratory scale development (Hair et al., 2010).

Descriptive Statistics and Correlations

Table 2.5 presents the descriptive statistics for each SRQ-IR subscale. Each subscale can range from 5-30, except *Rivalry* (5-20), as it comprises fewer items. Participants reported fairly high levels of *Warmth/Closeness* (M =24.29, SD = 4.60; *n* = 449), *Conflict Resolution* (M = 23.19, SD = 5.61; n = 445), and *Mentalising* (M = 22.28, SD = 5.39; *n* = 446) between their children. *Rivalry* (M = 10.26, SD = 4.64; n = 453) was near the midpoint of its possible range and *Antagonism* (M = 15.29, SD = 7.23; n = 445) and *Relational Aggression* (M = 12.37, SD = 6.69; *n* = 436) scores were the mid-point to lower end.

Table 2.5

Total Descriptive Statistics for SRQ-IR Subscales

Subscale	**Range**	**M**	**SD**
Antagonism	5-30	15.29	7.23
Warmth/Closeness	5-30	24.29	4.60
Relational Aggression	5-30	12.37	6.69
Conflict Resolution	5-30	23.19	5.61
Mentalising	5-30	22.28	5.39
Rivalry	5-20	10.26	4.64

Note. *M* = mean; *SD* = standard deviation.

Table 2.6 presents the descriptive statistics for Pre-school, Middle Childhood and Adolescent scores for each SRQ-IR subscales. For Preschool children, the highest scores were reported for *Warmth/Closeness* (*M* = 23.62, *SD* = 4.97) and *Conflict Resolution* (*M* = 21.81, *SD* = 5.92), suggesting strong positive sibling interactions, despite relatively high Antagonism (*M* = 17.84, *SD* = 7.00) and *Relational Aggression* (*M* = 14.90, *SD* = 7.64) levels. *Mentalising* (*M* = 21.13, *SD* = 5.67) and *Rivalry* (*M* = 11.53, *SD* = 4.66) were also moderately high.

In Middle Childhood, participants reported the highest levels of *Warmth/Closeness* (*M* = 24.41, *SD* = 4.33) and *Conflict Resolution* (*M* = 23.66, *SD* = 5.25) across all age groups. *Antagonism* (*M* = 13.75, *SD* = 6.61) and *Relational Aggression* (*M* = 11.02, *SD* = 5.53) were notably lower than in Preschool children. *Mentalising* remained high (*M* = 22.22, *SD* = 5.17), while *Rivalry* (*M* = 9.29, *SD* = 4.19) was the lowest across all age groups.

In Adolescence, scores remained relatively stable for *Warmth/Closeness* (*M* = 23.89, *SD* = 4.72) and *Conflict Resolution* (*M* = 23.04, *SD* = 5.53). *Antagonism* (*M* = 14.41, *SD* = 6.61) and *Relational Aggression* (*M* = 11.53, *SD* = 5.63) were moderate, reflecting some ongoing sibling conflict. *Mentalising* (*M* = 22.34, *SD* = 5.38) remained consistent with middle childhood levels, while *Rivalry* (M = 10.10, SD = 4.65) slightly increased compared to middle childhood but was still lower than preschool levels.

Table 2.6

Preschool, Middle Childhood and Adolescent Children Descriptive Statistics for SRQ-IR Subscales

		Preschool Children (1-5 years)		**Middle Childhood (6-11 years)**		**Adolescence (12-18 years)**	
Subscale	**Range**	**M**	**SD**	**M**	**SD**	**M**	**SD**

Antagonism	5-30	17.84	7.0	13.75	6.61	14.41	6.61
Warmth/Closeness	5-30	23.62	4.97	24.41	4.33	23.89	4.72
Relational Aggression	5-30	14.90	7.64	11.02	5.53	11.53	5.63
Conflict Resolution	5-30	21.81	5.92	23.66	5.25	23.04	5.53
Mentalising	5-30	21.13	5.67	22.22	5.17	22.34	5.38
Rivalry	5-20	11.53	4.66	9.29	4.19	10.10	4.65

Note. *M* = mean; *SD* = standard deviation.

Table 2.7 presents the descriptive statistics for Biological and Non-biological parents of siblings' scores for each SRQ-IR subscale. For Biological parents of siblings, the highest mean scores were reported for *Warmth/Closeness* (*M* = 24.41, *SD* = 4.53), followed by *Conflict Resolution* (*M* = 23.26, *SD* = 5.65) and *Mentalising* (*M* = 22.38, *SD* = 5.37). *Relational Aggression* had a mean score of *M* = 12.13 (*SD* = 6.61), while *Antagonism* was slightly lower at *M* = 12.67 (*SD* = 6.13). *Rivalry* (*M* = 10.10, SD = 4.65) was positioned near the midpoint of its possible range.

For the Non-biological parents of siblings (foster, adoptive, stepparent and guardians), the highest mean scores were similarly observed in *Warmth/Closeness* (*M* = 24.64, *SD* = 4.43), *Conflict Resolution* (*M* = 23.46, *SD* = 5.61), and *Mentalising* (*M* = 22.53, *SD* = 5.42). The mean score for *Antagonism* was *M* = 15.05 (*SD* = 7.34), while *Relational Aggression* had a mean of *M* = 11.81 (*SD* = 6.53). *Rivalry scored M* = 10.05 (*SD* = 4.69).

Table 2.7

Biological and Non-biological Parents of Siblings Descriptive Statistics for SRQ-IR Subscales

		Biological Parents		**Non-biological Parents**	
Subscale	**Range**	**M**	**SD**	**M**	**SD**
Antagonism	5-30	12.67	6.13	15.05	7.34

Warmth/Closeness	5-30	24.41	4.53	24.64	4.43
Relational Aggression	5-30	12.13	6.61	11.81	6.53
Conflict Resolution	5-30	23.26	5.65	23.46	5.61
Mentalising	5-30	22.38	5.37	22.53	5.42
Rivalry	5-20	10.10	4.65	10.05	4.69

Note. *M* = mean; *SD* = standard deviation.

Pearson correlations among the six sibling relationship factors are shown in Table 2.6. *Antagonism* was positively correlated with both *Relational Aggression* (r=.66, *p*<.01) and *Rivalry* (r=.60, *p*<.01), but negatively correlated with *Warmth/Closeness* (r=−.28, *p*<.01). *Warmth/Closeness* demonstrated moderate positive correlations with both *Conflict Resolution* (r=.53, *p*<.01) and *Mentalising* (r=.58, *p*<.01), reflecting that more positive aspects of sibling interactions tended to co-occur. Notably, *Mentalising* was weakly negatively correlated with *Rivalry* (r=−.10, *p*<.05).

Table 2.8

Pearson Correlations Between the SRQ-IR Subscales

Subscale	**Antagonism**	**Warmth/Closeness**	**Relational Aggression**	**Conflict Resolution**	**Mentalising**	**Rivalry**
Antagonism	1	-.28**	.66**	-.19**	-.22**	.60**
Warmth/Closeness		1	-.17**	.53**	.58**	-.12**
Relational Aggression			1	-.13**	-0.04	.68**
Conflict Resolution				1	.62**	-0.06
Mentalising					1	-.10*
Rivalry						1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

The data analysis process can be seen in Appendix 2.7.

Discussion

This study aimed to develop and validate a new measure of sibling relationships. In accordance with contemporary psychometric standards, it addressed limitations in existing tools by incorporating a broader range of sibling dynamics across developmental stages. The newly developed 34-item Sibling Relationship Quality-Informant Report (SRQ-IR) lays the foundation for future empirical research and validation studies.

Key Findings and Factor Structure

The EFA with PA identified a six-factor solution: *Antagonism, Warmth/Closeness, Relational Aggression, Conflict Resolution, Mentalising*, and *Rivalry.* This factor structure aligns with previously identified core dimensions of sibling relationships, such as *Warmth, Conflict and Rivalry,* while introducing new constructs, *Mentalising, Conflict Resolution* and *Relational Aggression.* Consistent with existing literature, *Antagonism* and *Rivalry* emerged as central elements of negative sibling interactions. *Antagonism* captured overt quarrels, teasing, and annoyance, while *Rivalry* was characterised by perceived inequalities and competition (e.g., "One thinks the other is better"). *Relational Aggression* emerged as a distinct dimension involving more indirect, socially manipulative behaviours (e.g., "One attempts to sabotage the other"). These findings highlight the multifaceted nature of sibling conflict, with both direct and indirect forms of aggression contributing to rivalry and tension.

The findings also indicated that these six factors form two distinct clusters with *Warmth/Closeness, Conflict Resolution, Mentalising* occurring together, whilst *Antagonism, Relational Aggression, Rivalry* formed a separate cluster. This pattern mirrors traditional conceptualisations of sibling relationship quality consisting of positive and negative clusters. However, this may oversimplify the complexities of sibling dynamics. The factors forming two overarching clusters could be considered instead as a pro-social set (*Warmth/Closeness, Conflict Resolution, Mentalising)* and a conflict-oriented set (*Antagonism, Relational Aggression, Rivalry*). These clusters align

with previous research that characterises sibling relationships as inherently ambivalent, comprising supportive and adversarial features (Kramer, 2010). It is proposed that an optimal balance of both supportive and antagonistic interactions between siblings may be key for healthy sibling relationships, although we currently lack the understanding of the ideal mixture of these proportions of such behaviours (Conger, Bryant & Brennom, 2004; Kramer & Bank, 2005). A CFA would be useful to test this hypothesised factor structure and positive and negative subscales, respectively, each being more closely related. A CFA could also test for invariance across developmental period.

This research introduced new constructs *Mentalising*, *Conflict Resolution* and *Relational Aggression*. Although the similar constructs of *Communication, Alienation* (SAI; Noel, Frances & Tilley, 2018), *Empathy*, *Aggression*, *Avoidance* (Xu et al., 2023) have been previously identified, the emergence of *Mentalising* highlights siblings' engagement in perspective-taking, empathetic understanding, and emotional insight, dimensions related to theory of mind, empathy, and psychological mindedness (Choi-Kain & Gunderson, 2008; Fonagy & Luyten, 2009; Bateman & Fonagy, 2019). Although some sibling scales have touched upon *Empathy* or *Communication* (Noel, Francis & Tilley, 2018; Xu et al., 2023), *Mentalising* as a distinct subdimension is less frequently explored, underscoring the SRQ-IR's novelty.

The construct *Conflict Resolution* parallels pro-social constructs of conflict management (Kramer, 2010), indicating that siblings use strategies to mitigate and repair conflicts. This aligns with research where constructive disputes facilitated negotiation, compromise, and emotional regulation (Katz, Kramer, & Gottman, 1992). Differentiating *Relational Aggression* from *Antagonism* clarifies the multifaceted nature of sibling conflict; while *Antagonism* refers to overt hostility (Volling & Elins, 1998), *Relational Aggression* includes socially manipulative behaviours such as exclusion and coercion, driven by anger and causing emotional harm (Crick & Grotpeter, 1995; Crick, 1995; Galen & Underwood, 1997). The separation of these factors highlights the complexity of sibling conflict, suggesting siblings engage in both confrontations and subtler, socially driven tactics. The absence of a significant correlation between

Mentalising and *Relational Aggression* suggests that relationally aggressive siblings may still have social insight but use it manipulatively rather than pro-socially.

Theoretical and Practical Implications

The SRQ-IR's multidimensional structure suggests that sibling relationships may not be adequately captured by a simple positive–negative dichotomy (Kramer, 2010). Instead, the results indicate that elements of conflict and cooperation often coexist, with the capacity to transition from disagreement to mutual support potentially playing a key role in healthy development. Theoretically, the measure's ability to capture shifts between warmth and hostility aligns with family systems theory and attachment perspectives, implying that complex sibling dynamics can nurture and challenge social development (Whiteman, McHale, & Soli, 2011).

Furthermore, the emergence of a distinct *Mentalising* factor points to its potential importance, suggesting that secure sibling bonds might foster the skills necessary to interpret each other's emotions and intentions (Fonagy et al., 2002; Fonagy, Luyten & Strathearn, 2011; Luyten & Fonagy, 2015). Although these findings are preliminary, the SRQ-IR may help clinicians, social workers, and educators identify family strengths, such as high *Warmth/Closeness* and effective *Conflict Resolution*, as well as potential vulnerabilities like elevated *Antagonism* and *Rivalry.* The SRQ-IR's *Mentalising* and *Relational Aggression* subscales could also provide a foundation for targeted interventions, for example, by emphasising perspective-taking skills when mentalising scores are low or by enhancing conflict resolution strategies when antagonism is high (Kramer, 2014)

Strengths and Limitations

The SRQ-IR extends the field of sibling relationships and its measurement in a number of ways. Firstly, it is applicable across a broader developmental range, as parents of children from diverse age groups were included. Another key strength is the content validity ensured through consultations with parents, clinicians, and researchers during item generation. This process helped produce items reflecting real-life sibling dynamics. Further, the structural validity (demonstrated through EFA and parallel analysis) and

strong internal consistency (Cronbach's alphas ranging from .82 to .93) suggest that the measure reliably captures distinct constructs. The global sample, which included participants from various ethnic backgrounds and non-traditional family structures, responds to calls for more inclusive research on siblings.

Despite these strengths, several limitations warrant consideration. EFA is useful for initial scale development and identifying potential factors, but CFA is necessary to would further validate and refine the factor structure. Second, despite our best efforts to reach a more diverse sample, the sample still skewed toward White, relatively well-educated parents, limiting broader generalisability. Third, the measure relied solely on parent/guardian reports, which may not fully capture the nuances of sibling interactions or the children's perspectives.

Future Directions

Building on these findings, future research could conduct a CFA on larger, more diverse samples to confirm the factor structure and improve measurement precision. Assuming the factor structure is stable, the SRQ-IR could inform the development of these measures. Extending participant diversity in future validation studies is critical to understanding how the SRQ-IR performs across cultural, socioeconomic and family contexts. The reliance on parent/guardian informant reports introduces potential biases; therefore, incorporating multi-informant approaches, including self-reports from siblings and observational data, would provide a more comprehensive assessment. Finally, longitudinal studies are needed to examine the stability of sibling dynamics over time and their responses to interventions.

Conclusion

The SRQ-IR provides a promising multidimensional framework that captures both pro-social and conflict-oriented aspects of sibling dynamics. This new measure can advance the field's understanding of sibling relationships by incorporating constructs such as *Mentalising*, *Conflict Resolution* and *Relational Aggression* alongside traditional *Warmth*, *Conflict* and *Rivalry* constructs. Although further research, including CFA, multi-informant designs, and diverse samples is necessary, these initial

findings underscore the value of adopting a comprehensive, contemporary, and inclusive perspective on sibling relationship quality.

References

Adamis, D., Tsamparl, A., & Talanti, K. (2017). Psychometric analysis of the Greek version of the Sibling Relationship Questionnaire. *Psychology: The Journal of the Hellenic Psychological Society, 22(1),* 1–13. https://doi.org/10.12681/psy_hps.23175

Bateman, A., & Fonagy, P. (2019). *Handbook of mentalizing in mental health practice* (2nd ed.). American Psychiatric Press.

Ben-Arieh, A., Casas, F., Frønes, I., & Korbin, J. E. (2014). Handbook of Child Well-Being. Springer eBooks. https://doi.org/10.1007/978-90-481-9063-8

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

Buist, K. L., Deković, M., & Prinzie, P. (2013). Sibling relationship quality and psychopathology of children and adolescents: a meta-analysis. *Clinical Psychology Review*, *33*(1), 97–106. https://doi.org/10.1016/j.cpr.2012.10.007

Buhrmester, D., & Furman, W. (1990). Perceptions of sibling relationships during middle childhood and adolescence. *Child Development, 61*(5), 1387–1398. https://doi.org/10.2307/1130750

Caspi, J. (2011). Future directions for sibling research, practice, and theory. In J. Caspi (Ed.), *Sibling development: Implications for mental health practitioners* (pp. 377–390). Springer Publishing Company.

Choi-Kain, L. W., & Gunderson, J. G. (2008). Mentalization: ontogeny, assessment, and application in the treatment of borderline personality disorder. *The American Journal of Psychiatry*, *165*(9), 1127–1135. https://doi.org/10.1176/appi.ajp.2008.07081360

Cicirelli, V. G. (1995). Sibling relationships across the life span. Plenum Press. https://doi.org/10.1007/978-1-4757-6509-0

Clark, L. & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. Psychological Assessment, 7, 309-319.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*(7), 1–9.

Crick, N. R. (1995). Relational aggression: The role of intent attributions, feelings of distress, and provocation type. *Development and Psychopathology, 7*(2), 313–322. https://doi.org/10.1017/S0954579400006520

Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child development*, *66*(3), 710–722. https://doi.org/10.1111/j.1467-8624.1995.tb00900.x

Conger, K. J., Bryant, C. M., & Brennom, J. M. (2004). The changing nature of adolescent sibling relationships. In R. D. Conger, F. O. Lorenz, & K.A.S. Wickrama (Eds.), *Continuity and change in family relations: Theory, methods, and empirical findings* (pp. 319–344). Mahwah, NJ: Erlbaum.

Derkman, M. M. S., Scholte, R. H. J., Van der Veld, W. M., & Engels, R. C. M. E. (2010). Factorial and construct validity of the Sibling Relationship Questionnaire. *European Journal of Psychological Assessment, 26*(4), 277–283. https://doi.org/10.1027/1015-5759/a000037

European Union. (2018). *General Data Protection Regulation (GDPR) (EU) 2016/679*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

Field, A. (2013) Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock “N” Roll, 4th Edition, Sage, Los Angeles, London, New Delhi.

Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect Regulation, Mentalization and the Development of the Self.* New York: Other Press.

Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and psychopathology*, *21*(4), 1355–1381. https://doi.org/10.1017/S0954579409990198

Fonagy, P., Luyten, P., & Strathearn, L. (2011). Borderline personality disorder, mentalization, and the neurobiology of attachment. *Infant mental health journal*, *32*(1), 47–69. https://doi.org/10.1002/imhj.20283

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Furman, W., & Buhrmester, D. (1985). Children's Perceptions of the Qualities of Sibling Relationships. Child Development, 56(2), 448–461. https://doi.org/10.2307/1129733

Galen, B. R., & Underwood, M. K. (1997). A developmental investigation of social aggression among children. *Developmental Psychology, 33*(4), 589–600. https://doi.org/10.1037/0012-1649.33.4.589

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). Multivariate Data Analysis (7th ed.). Pearson.

Hetherington, E. M. (1989). Coping with family transitions: Winners, losers, and survivors. *Child Development, 60*(1), 1–14. https://doi.org/10.2307/1131066

Holmes, M. R., Bender, A. E., O'Donnell, K. A., Miller, E. K., & Conard, I. T. (2024). Illuminating the landscape of sibling relationship quality: An evidence and gap map. *Child Development*. https://doi.org/10.1111/cdev.14065

Hughes, C., McHarg, G., & White, N. (2018). Sibling influences on prosocial behavior. *Current Opinion in Psychology*, *20*, 96–101. https://doi.org/10.1016/j.copsyc.2017.08.015

IBM Corp. (2023). *IBM SPSS Statistics for Mac, Version 28.0*. Armonk, NY: IBM Corp

Jiang, M., Cao, X., Huang, Q., Wu, S., & Chen, X. (2022). Exploring the structure of sibling relationships among preschool children in China and developing a questionnaire. *Frontiers in Psychology, 12*, Article 745165. https://doi.org/10.3389/fpsyg.2021.745165

Katz, L. F., Kramer, L., & Gottman, J. M. (1992). Conflict and emotions in marital, sibling, and peer relationships. In C. U. Shantz & W. W. Hartup (Eds.), *Conflict in child and adolescent development* (pp. 122–149). Cambridge University Press.

Kramer, L. (2010). The essential ingredients of successful sibling re- lationships: An emerging, the essential ingredients of successful sibling relationships: An emerging framework for advancing theory and practice framework for advancing theory and prac- tice. *Child Development Perspectives*, *4*, 80–86. https://doi.org/10. 1111/j.1750-8606.2010.00122.x

Kramer, L. (2014). Learning emotional understanding and emotion regulation through sibling interaction. *Early Education and Development*, *25*, 160–184. https://doi.org/10.1080/10409289.2014. 838824

Kramer, L., & Bank, L. (2005). Sibling relationship contributions to individual and family well-being: introduction to the special issue. *Journal of family psychology : JFP : journal of the Division of Family Psychology of the American Psychological Association (Division 43)*, *19*(4), 483–485. https://doi.org/10.1037/0893-3200.19.4.483

Kramer, L., & Baron, L. A. (1995). Parental perceptions of children's sibling relationships. *Family Relations: An Interdisciplinary Journal of Applied Family Studies, 44*(1), 95–103. https://doi.org/10.2307/584746

López-Fernández, G., Gómez-Benito, J., Kramer, L., & Barrios, M. (2022). Spanish validation of the Parental Expectations and Perceptions of Children's Sibling Relationships Questionnaire. *Family Relations.*https://doi.org/10.1111/fare.12699

Luyten, P., & Fonagy, P. (2015). The neurobiology of mentalizing. *Personality disorders*, *6*(4), 366–379. https://doi.org/10.1037/per0000117

McHale, S. M., Updegraff, K. A., & Whiteman, S. D. (2012). Sibling relationships and influences in childhood and adolescence. *Journal of Marriage and Family*, 74(5), 913–930. https://doi.org/10.1111/j.1741-3737.2012.01011.x

Milesky, A., & Levitt, M. (2005). Sibling support in early adolescence: Buffering and compensation across relationships. The European Journal of Developmental Psychology, 2(3), 299–320. https://doi.org/10.1080/17405620544000048

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psychology: Research and Review, 30*(1), 3. https://doi.org/10.1186/s41155-016-0057-1

Noel, V. A., Francis, S. E., & Tilley, M. A. (2018). An adapted measure of sibling attachment: Factor structure and internal consistency of the Sibling Attachment Inventory in youth. *Child Psychiatry & Human Development, 49*(2), 217–224. https://doi.org/10.1007/s10578-017-0742-z

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32,* 396-402.

Persram, R. J., Howe, N., & Bukowski, W. M. (2022). Constructing and validating a measure of adolescents' perceptions of sibling trust. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 54*(1), 73–78. https://doi.org/10.1037/cbs0000281

Stocker, C. M., & McHale, S. M. (1992). The nature and family correlates of preadolescents' perceptions of their sibling relationships. *Journal of Social and Personal Relationships, 9*(2), 179–195. https://doi.org/10.1177/0265407592092003

Swift, E. E., Taylor, H. G., Kaugars, A. S., Drotar, D., Yeates, K. O., Wade, S. L., & Stancin, T. (2003). Sibling relationships and behavior after pediatric traumatic brain injury. *Journal of Developmental and Behavioral Pediatrics, 24*(1), 24–31. https://doi.org/10.1097/00004703-200302000-00004

Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and Solutions in Human Assessment: Honoring Douglas Jackson at seventy* (pp. 41–71). Boston, MA: Kluwer. http://dx.doi.org/10.1007/978-1-4615-4397-8_3

Volling, B. L., & Elins, J. L. (1998). Family relationships and children's emotional adjustment as correlates of maternal and paternal differential treatment: a replication with toddler and preschool siblings. *Child development*, *69*(6), 1640–1656

Wang, Y. Z., Li, Y., Liu, T., Zhao, J., & Li, Y. (2021). Understanding relationships within cultural contexts: Developing an early childhood sibling relationship questionnaire in China. *Family Relations.* https://doi.org/10.1111/fare.12586

Whiteman, S. D., McHale, S. M., & Soli, A. (2011). Theoretical Perspectives on Sibling Relationships. *Journal of Family Theory & Review*, *3*(2), 124–139. https://doi.org/10.1111/j.1756-2589.2011.00087.x

Xu, H., Wang, Z., Gao, X., Wang, X., & Wu, Q. (2023). Psychometric validation of the sibling inventory of behavior in three- to six-year-old Chinese children. *Frontiers in Psychology, 14,* 1124518. https://doi.org/10.3389/fpsyg.2023.1124518

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442. http://dx.doi.org/10.1037/0033-2909.99.3.432

AI Software

Grammarly. (2025). *Grammarly AI* (Version 2.0). Grammarly, Inc. https://www.grammarly.com

The Grammarly Premium AI tool was used to check grammar and style to help improve the overall flow of writing for this piece of work. This included features such as review checking – which provides feedback and suggestions on correctness, clarity, engagement and writing delivery.

Appendices

Appendix 1.1 – Search Strategy

Database	Search Terms	Filters Applied
PsycINFO	1. ("sibling relationship*" OR "sibling bond*" OR "sibling quality" OR "sibling attachment" OR "sibling closeness" OR "sibling support" OR "sibling conflict" OR "sibling rivalry") 2. ("measure*" OR "instrument*" OR "tool*" OR "questionnaire*" OR "inventory" OR "rating scale*" OR "scale*" OR "assessment") 3. ("psychometric*" OR "content validity" OR "face validity" OR "construct validity" OR "criterion validity" OR "internal consistency" OR "reliability" OR "test-retest" OR "factor analysis" OR "exploratory factor analysis" OR "confirmatory factor analysis" OR "EFA" OR "CFA")	Age: 2-18 years Language: English Peer-reviewed: Yes Document Type: Exclude dissertations, book chapters, conference proceedings
APA PsycTests	1. ("sibling relationship*" OR "sibling bond*" OR "sibling quality" OR "sibling attachment" OR "sibling closeness" OR "sibling support" OR "sibling conflict" OR "sibling rivalry") 2. ("test" OR "instrument" OR "tool" OR "measure" OR "questionnaire" OR "rating scale" OR "scale" OR "survey" OR "inventory" OR "assessment") 3. (Optional) ("psychometric*" OR "content validity" OR "internal consistency" OR "test-retest" OR "validity" OR "reliability" OR "factor analysis" OR "CFA" OR "EFA" OR "construct validity")	Age: 2-18 years Language: English Test Purpose: Development, validation, psychometric evaluation
CINAHL	1. ("sibling relationship*" OR "sibling bond*" OR "sibling quality" OR "sibling attachment" OR "sibling closeness" OR "sibling support" OR "sibling conflict" OR "sibling rivalry") 2. ("measure*" OR "instrument*" OR "tool*" OR "questionnaire*" OR "rating scale*" OR "scale*" OR "survey" OR "inventory" OR "assessment") 3. ("psychometric*" OR "content validity" OR "internal consistency" OR "test-retest" OR "validity" OR "reliability" OR "factor analysis" OR "CFA" OR "EFA" OR "construct validity")	Age: 2-18 years Language: English Publication Type: Peer-reviewed journals Document Type: Exclude grey literature

MEDLINE	1. ("sibling relationship*" OR "sibling bond*" OR "sibling quality" OR "sibling attachment" OR "sibling closeness" OR "sibling support" OR "sibling conflict" OR "sibling rivalry") 2. ("measure*" OR "instrument*" OR "tool*" OR "questionnaire*" OR "rating scale*" OR "scale*" OR "survey" OR "inventory" OR "assessment") 3. ("psychometric*" OR "content validity" OR "internal consistency" OR "test-retest" OR "validity" OR "reliability" OR "factor analysis" OR "CFA" OR "EFA" OR "construct validity")	Age: 2-18 years Language: English Publication Type: Peer-reviewed journal articles Document Type: Exclude grey literature (dissertations, theses, etc.)
Ovid (Embase + HaPI)	1. ("sibling relationship*" OR "sibling bond*" OR "sibling quality" OR "sibling attachment" OR "sibling closeness" OR "sibling support" OR "sibling conflict" OR "sibling rivalry") 2. ("measure*" OR "instrument*" OR "tool*" OR "questionnaire*" OR "rating scale*" OR "scale*" OR "survey" OR "inventory" OR "assessment") 3. ("psychometric*" OR "content validity" OR "internal consistency" OR "test-retest" OR "validity" OR "reliability" OR "factor analysis" OR "CFA" OR "EFA" OR "construct validity")	Age: 2-18 years Language: English Publication Type: Peer-reviewed journal articles Document Type: Exclude grey literature (dissertations, theses, etc.)

Appendix 1.2 – Data Extraction Checklist

1. Study Identification	• Author(s) • Year of Publication • Journal Name • Country
2. Study Characteristics	• Study Design • Purpose of Study: • Recruitment Base • Eligibility Criteria
3. Participant Information	• Sample Size • Age Range of Participants • Ethnicity (if reported) • Other Relevant Demographics • Population Type • Study Setting
4. Sibling Relationship Measure Information	• Name of Measure • Type of Measure • Purpose of Measure • Underlying Factors/Constructs Measured • Scoring Method • Response Options • Child or Parent Reported • Intended Context of Use

5. Psychometric Properties Assessed Using COSMIN	• Content Validity
	• Structural Validity
	• Internal Consistency
	• Cross-Cultural Validity
	• Reliability
	• Measurement Error
	• Criterion Validity:
	• Hypothesis Testing for Construct Validity:
	  o Convergent Validity
	  o Discriminant Validity
	• Responsiveness

Appendix 1.3 – COSMIN Risk of Bias Quality Appraisal Tool

Scope of the PROM	
1 Is a clear description provided of the construct to be measured?	○ Construct clearly described ○ Construct not clearly described
2 Is the origin of the construct clear: was a theory, conceptual framework or disease model used or clear rationale provided to define the construct to be measured?	○ Origin of the construct clear ○ Origin of the construct not clear
3 Is a clear description provided of the target population for which the PROM was developed?	○ Target population clearly described ○ Target population not clearly described
4 Is a clear description provided of the context of use	○ Context of use clearly described ○ Context of use not clearly described
5 What is the measurement model on which the PROM is based?	○ Reflective model ○ Formative model[1] ○ unclear

[1] If the scale is not based on a reflective model, unidimensionality or structural validity is not relevant.

Does the study concern unidimensionality or structural validity? [2]	○ unidimensionality ○ structural validity

Box 1. PROM development						
1a. Concept elicitation study (relevance and comprehensiveness)		**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
1	Was the concept elicitation study performed in a sample representing the target population for which the PROM was developed?	Study performed in a sample representing the target population	Assumable that the study was performed in a sample representing the target population, but not clearly described	Doubtful whether the study was performed in a sample representing the target population	Study not performed in a sample representing the target population	
2	Was an appropriate qualitative data collection method used to identify relevant items for a new PROM?	Widely recognized or well justified qualitative method used, suitable for the construct and study population	Assumable that the qualitative method was appropriate and suitable for the construct and study population, but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was suitable for the construct and study population	Method used not appropriate or not suitable for the construct or study population	
3	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained or group moderators /interviewers not trained and no experience		NA

4	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate or no guide		NA
5	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings of interviews were recorded and transcribed verbatim or recordings not transcribed verbatim or only notes were made during the group meetings/ interviews	No recording and no notes	NA
6	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used or doubtful whether the approach was appropriate	Approach not appropriate	
7	Was at least part of the data coded independently?	At least 50% of the data was coded by at least two researchers independently	11-49% of the data was coded by at least two researchers independently	Doubtful if two researchers were involved in the coding or only 1-10% of the data was coded by at least two researchers independently	Only one researcher was involved in coding or no coding	Not applicable
8	Was data collection continued until saturation was reached?	Evidence provided that saturation was reached	Assumable that saturation was reached	Doubtful whether saturation was reached	Evidence suggests that saturation was not reached	NA
9	For quantitative studies (surveys): was the sample size appropriate?	≥100	50-99	30-49	<30	NA

Other						
10	Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

1b. Pilot study (Cognitive interview study or other pilot test) (comprehensibility)					
	very good	**adequate**	**doubtful**	**inadequate**	**NA**
Comprehensibility					
11 Was the pilot study performed in a sample representing the target population for which the PROM was developed?	Study performed in a sample representing the target population AND Sample was not included in the concept elicitation study	Assumable that the study was performed in a sample representing the target population but not clearly described, or sample was included in the concept elicitation study	Doubtful whether the study was performed in a sample representing the target population	Study not performed in a sample representing the target population	
12 Was the comprehensibility assessed of the PROM instructions, items, response options, and recall period?	Comprehensibility of the PROM instructions, items, response options, and recall period was assessed		Not clear if patients were asked about the comprehensibility of all items, response options, instructions, and recall period OR patients were not asked about the comprehensibility of the PROM instructions or the recall period	Patients were not asked about the comprehensibility of all items and response options	

13	Were all items tested in their final form?	All items were tested in their final form	Assumable that all items were tested in their final form, but not clearly described	Not clear if all items were tested in their final form	Items were not tested in their final form or items were not re-tested after substantial adjustments	
14	Was an appropriate qualitative method used?	Widely recognized or well justified qualitative method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used or doubtful whether the method was appropriate	Method used not appropriate	
15	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
16	Were skilled interviewers used?	Skilled interviewers used	Interviewers had limited experience or were trained specifically for the study	Not clear if interviewers were trained OR interviewers were not and had no experience		NA
17	Were the interviews based on an appropriate interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic or interview guide was appropriate OR no guide		NA
18	Were the interviews recorded and transcribed verbatim?	All interviews were recorded and transcribed verbatim	Assumable that all interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all interviews were recorded and transcribed verbatim OR recordings not transcribed verbatim OR only notes were made during the interviews	No recording and no notes	NA

19	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
20	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
21	Were problems regarding the comprehensibility of the PROM instructions, items, response options, and recall period appropriately addressed by adapting the PROM?	No problems found OR problems appropriately addressed and PROM was adapted and re-tested if necessary	Assumable that there were no problems or that problems were appropriately addressed, but not clearly described	Not clear if there were problems OR doubtful if problems were appropriately addressed	Problems not appropriately addressed OR PROM was adapted but items were not re-tested after substantial adjustments	NA
Other						
22	Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 2. Content validity

2a. Asking patients about relevance

Design requirements		very good	adequate	doubtful	inadequate	NA
1	Was an appropriate method used to ask patients whether each item is relevant for their experience with the condition?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used or assumable that the method was appropriate but not clearly described	Not clear if patients were asked whether each item is relevant OR doubtful whether the method was appropriate	Method used not appropriate OR patients not asked about the relevance of all items	
2	Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
3	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained OR group moderators /interviewers were not trained and had no experience		NA
4	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used OR doubtful if topic or interview guide was appropriate OR no guide		NA

5	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim OR recordings not transcribed verbatim OR only notes were made during the group meetings/ interviews	No recording and no notes	NA
Analyses						
6	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
7	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other						
8	Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

2b Asking patients about comprehensiveness						
Design requirements		**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
9	Was an appropriate method used for assessing the comprehensiveness of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used OR assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
10	Was the PROM tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
11	Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained OR group moderators /interviewers were not trained and had no experience		NA
12	Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used OR doubtful if topic or interview guide was appropriate OR no guide		NA

13	Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim OR recordings not transcribed verbatim OR only notes were made during the group meetings/ interviews	No recording and no notes	NA
Analyses						
14	Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
15	Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other						
16	Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

2c Asking patients about comprehensibility					
Design requirements	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
17 Was an appropriate qualitative method used for assessing the comprehensibility of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified qualitative method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used OR doubtful whether the method was appropriate OR not clear if patients were asked about the comprehensibility of the items, response options and recall period OR patients not asked about the comprehensibility of the PROM instructions or recall period	Method used not appropriate OR patients not asked about the comprehensibility of the items, response options, and recall period	
18 Was each item tested in an appropriate number of patients? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
19 Were skilled group moderators/interviewers used?	Skilled group moderators/ interviewers used	Group moderators /interviewers had limited experience or were trained specifically for the study	Not clear if group moderators /interviewers were trained OR group moderators /interviewers were not trained and had no experience		
20 Were the group meetings or interviews based on an appropriate topic or interview guide?	Appropriate topic or interview guide	Assumable that the topic or interview guide was appropriate, but not clearly described	Not clear if a topic guide was used or doubtful if topic OR interview guide was appropriate OR no guide		NA
21 Were the group meetings or interviews recorded and transcribed verbatim?	All group meetings or interviews were recorded and transcribed verbatim	Assumable that all group meetings or interviews were recorded and transcribed verbatim, but not clearly described	Not clear if all group meetings or interviews were recorded and transcribed verbatim OR recordings not transcribed verbatim OR only notes were made during the group meetings/ interviews	No recording and no notes	NA
Analyses					
22 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
23 Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other					
24 Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

2d. Asking professionals about relevance					
Design requirements	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
25 Was an appropriate method used to ask professionals whether each item is relevant for the construct of interest?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used OR assumable that the method was appropriate but not clearly described	Not clear if professionals were asked whether each item is relevant OR doubtful whether the method was appropriate	Method used not appropriate OR professionals not asked about the relevance of all items	
26 Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included OR relevant professionals were not included		
27 Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
Analyses					
28 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	

29 Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other					
30 Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

2e. Asking professionals about comprehensiveness					
Design requirement	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
31 Was an appropriate method used for assessing the comprehensiveness of the PROM?	Widely recognized or well justified method used	Only quantitative (survey) method(s) used OR assumable that the method was appropriate but not clearly described	Doubtful whether the method was appropriate	Method used not appropriate	
32 Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included OR relevant professionals were not included		

33 Was the PROM tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		
Analyses					
34 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
35 Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other					
36 Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

2f. Asking professionals about comprehensibility					
Design requirement	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
37 Was an appropriate method used for assessing the comprehensibility of the PROM instructions, items, response options, and recall period?	Widely recognized or well justified method used	Assumable that the method was appropriate but not clearly described	Only quantitative (survey) method(s) used OR doubtful whether the method was appropriate OR not clear if professionals were asked about the comprehensibility of all items, response options instructions, and recall period OR professionals were not asked about the comprehensibility of the PROM instructions or the recall period	Method used not appropriate OR professionals not asked about the comprehensibility of all items and response options	
38 Were professionals from all relevant disciplines included?	Professionals from all required disciplines were included	Assumable that professionals from all required disciplines were included, but not clearly described	Doubtful whether professionals from all required disciplines were included OR relevant professionals were not included		
39 Was each item tested in an appropriate number of professionals? For qualitative studies For quantitative (survey) studies	 ≥7 ≥50	 4-6 ≥30	 <4 or not clear <30 or not clear		

Analyses					
40 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	
41 Were at least two researchers involved in the analysis?	At least two researchers involved in the analysis	Assumable that at least two researchers were involved in the analysis, but not clearly described	Not clear if two researchers were included in the analysis OR only one researcher involved in the analysis		
Other					
42 Were there any other important flaws in the design or methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 3. Structural validity

Statistical methods	very good	adequate	doubtful	inadequate	NA
1 For CTT: Was exploratory or confirmatory factor analysis performed?	Confirmatory factor analysis performed	Exploratory factor analysis performed	Only PCA was performed	No exploratory or confirmatory factor analysis performed	NA
2 For IRT/Rasch: does the chosen model fit to the research question?	Chosen model fits well to the research question	Assumable that the chosen model fits well to the research question	Doubtful if the chosen model fits well to the research question	Chosen model does not fit to the research question	NA
3 Was the sample size included in the analysis adequate?	FA: 7 times the number of items in the tested model and ≥100	FA: at least 5 times the number of items in the tested model and ≥100; OR at least 6 times number of items in the tested model but <100	FA: 5 times the number of items in the tested model but <100	FA: < 5 times the number of items in the tested model	
	Rasch/1PL models: ≥ 200 subjects	Rasch/1PL models: 100-199 subjects	Rasch/1PL models: 50-99 subjects	Rasch/1PL models: < 50 subjects	
	2PL parametric IRT models OR Mokken scale analysis: ≥ 1000 subjects	2PL parametric IRT models OR Mokken scale analysis: 500-999 subjects	2PL parametric IRT models OR Mokken scale analysis: 250-499 subjects	2PL parametric IRT models OR Mokken scale analysis: < 250 subjects	
Other					
4 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws (e.g. rotation method not described)	Other important methodological flaws (e.g. inappropriate rotation method)	

Box 4. Internal consistency

Statistical methods	very good	adequate	doubtful	inadequate	NA
1 For continuous scores: Was Cronbach's alpha or omega calculated?	Cronbach's alpha, or Omega calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated	NA
2 For dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated	NA
3 For IRT-based scores: Was standard error of the theta (SE (θ)) or reliability coefficient of estimated latent trait value (index of (subject or item) separation) calculated?	SE(θ) or reliability coefficient calculated			SE(θ) or reliability coefficient NOT calculated	NA
Other					
4 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 5. Cross-cultural validity\Measurement invariance

Design requirements	very good	adequate	doubtful	inadequate	NA
1 Were the samples similar for relevant characteristics except for the group variable?	Evidence provided that samples were similar for relevant characteristics except group variable	Stated (but no evidence provided) that samples were similar for relevant characteristics except group variable	Unclear whether samples were similar for relevant characteristics except group variable	Samples were NOT similar for relevant characteristics except group variable	
Statistical methods					
2 Was an appropriate approach used to analyse the data?	A widely recognized or well justified approach was used	Assumable that the approach was appropriate, but not clearly described	Not clear what approach was used OR doubtful whether the approach was appropriate	Approach not appropriate	NA
3 Was the sample size included in the analysis adequate?	MGCFA: 7 times the number of items in the model and ≥100	5 times the number of items in the model and ≥100; OR 5-7 times the number of items in the model but <100	5 times the number of items in the model but <100	<5 times the number of items in the model	
	Regression analyses or IRT/Rasch based analyses: 200 subjects per group	150 subjects per group	100 subjects per group	< 100 subjects per group	
Other					
4 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 6. Reliability

Design requirements	very good	adequate	doubtful	inadequate	NA
1 Were patients stable on the construct to be measured in the time between the repeated measurements?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2 Was the time interval between the repeated measurements appropriate?	Time interval appropriate		Doubtful if time interval was appropriate OR time interval was not stated	Time interval NOT appropriate	
3 Were the measurement conditions similar for the repeated measurements – except for the condition being evaluated?	Measurement conditions were similar (evidence provided)	Assumable that measurement conditions were similar	Unclear if measurement conditions were similar	Measurement conditions were NOT similar	
Statistical methods					
4 For continuous scores: Was the appropriate intraclass correlation coefficient (ICC) calculated?	$ICC_{agreement}$ was or can be calculated	($ICC_{consistency}$, Pearson or Spearman correlation coefficient was calculated, OR the ICC model or formula was not described) WITH evidence provided that no systematic change between measurements has occurred OR $ICC_{one\text{-}way}$ was calculated	($ICC_{consistency}$, Pearson or Spearman correlation coefficient was calculated OR the ICC model or formula was not described) WITHOUT evidence provided that no systematic change between measurements has occurred	($ICC_{consistency}$, Pearson or Spearman correlation coefficient was calculated OR the ICC model or formula was not described) WITH evidence provided that a systematic change between measurements has occurred	NA
5 For dichotomous scores: was kappa calculated?	Kappa calculated				NA
6 For nominal scores: was an unweighted kappa calculated?	Unweighted kappa calculated				NA
7 For ordinal scores: was a weighted kappa calculated?	Weighted kappa calculated and the weighting scheme was described	Kappa calculated, but weighting scheme not described	Unweighted Kappa calculated or unclear if weighting was applied		NA
Other					
8 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 7. Measurement error

Design requirements	very good	adequate	doubtful	Inadequate	NA
1 Were patients stable on the construct to be measured in the time between the repeated measurements?	Evidence provided that patients were stable	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable	
2 Was the time interval between the repeated measurements appropriate?	Time interval appropriate		Doubtful if time interval was appropriate OR time interval was not stated	Time interval NOT appropriate	
3 Were the measurement conditions similar for the measurements – except for the condition being evaluated as a source of variation?	Measurement conditions were similar (evidence provided)	Assumable that measurement conditions were similar	Unclear if measurement conditions were similar	Measurement conditions were NOT similar	

Statistical methods						
4	For continuous scores: was the Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) or Limits of Agreement (LoA) calculated?	$SEM_{agreement}$, $SEM_{one-way}$, $SDC_{agreement}$, or $SDC_{one-way}$ was or could be calculated	($SEM_{consistency}$ or $SDC_{consistency}$ or LoA was calculated OR the SEM/SDC model or formula was not described) WITH evidence provided that no systematic change between the measurement has occurred	($SEM_{consistency}$, $SDC_{consistency}$, or LoA was calculated) WITHOUT evidence provided that no systematic change between measurements has occurred	SEM calculated based on Cronbach's alpha or SD from another population OR ($SEM_{consistency}$ $SDC_{consistency}$ or LoA calculated) WITH evidence provided that systematic change has occurred	NA
5	For dichotomous/nominal/ordinal scores: was the percentage (positive and negative) agreement calculated?	% positive and negative agreement calculated	% agreement calculated			NA
Other						
6	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 8. Criterion validity

		very good	adequate	doubtful	inadequate	NA
Statistical methods						
1	For continuous scores: were correlations, or the AUC calculated?	Correlations or AUC calculated				NA
2	For dichotomous scores: were sensitivity and specificity determined?	Sensitivity and specificity calculated				NA
Other						
3	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 9. Hypotheses testing for construct validity

9a. Comparison with other outcome measurement instruments (convergent validity)

		very good	adequate	doubtful	inadequate	NA
Design requirements						
1	Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
2	Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s), OR evidence for insufficient measurement properties of the comparator instrument(s)	
Statistical methods						
3	Were statistical methods adequate for the comparisons made?	Statistical methods applied were appropriate	Assumable that statistical methods were appropriate	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate	
Other						
4	Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

9b. Comparison between subgroups (discriminative or known-groups validity)

Design requirements	very good	adequate	doubtful	inadequate	NA
5 Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor of no description of the important characteristics of the subgroups		
Statistical methods					
6 Were statistical methods appropriate for the subgroups being compared?	Statistical methods applied were appropriate	Assumable that statistical methods were appropriate	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate	
Other					
7 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Box 10. Responsiveness

10a. Criterion approach (i.e. comparison to a gold standard)

	very good	adequate	doubtful	inadequate	NA
Statistical methods					
1 For continuous scores: were correlations between change scores, or the AUC calculated?	Correlations or AUC calculated				NA
2 For dichotomous scales: were sensitivity and specificity (changed versus not changed) determined?	Sensitivity and specificity calculated				NA
Other					
3 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10b. Construct approach (i.e. hypotheses testing; comparison with other outcome measurement instruments)

Design requirements	very good	adequate	doubtful	inadequate	NA
4 Is it clear what the comparator instrument(s) measure(s)?	Constructs measured by the comparator instrument(s) is clear			Constructs measured by the comparator instrument(s) is not clear	
5 Were the measurement properties of the comparator instrument(s) sufficient?	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties of the comparator instrument(s) in any study population	NO information on the measurement properties of the comparator instrument(s) OR evidence of poor quality of comparator instrument(s)	
Statistical methods					
6 Were statistical methods appropriate for the comparisons being made?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate	
Other					
7 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10c. Construct approach: (i.e. hypotheses testing: comparison between subgroups)					
Design requirements	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
8 Was an adequate description provided of important characteristics of the subgroups?	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor or no description of the important characteristics of the subgroups		
Statistical methods					
9 Were statistical methods appropriate for the subgroups being compared?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate	Statistical method applied NOT optimal	Statistical method applied NOT appropriate	
Other					
10 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

10d. Construct approach: (i.e. hypotheses testing: before and after intervention)					
Design requirements	**very good**	**adequate**	**doubtful**	**inadequate**	**NA**
11 Was an adequate description provided of the intervention given?	Adequate description of the intervention		Poor description of the intervention	NO description of the intervention	
Statistical methods					
12 Were statistical methods appropriate for the before-after comparison being made?	Statistical methods applied appropriate	Assumable that statistical methods were appropriate	Statistical methods applied NOT optimal	Statistical methods applied NOT appropriate	
Other					
13 Were there any other important flaws in the design or statistical methods of the study?	No other important methodological flaws		Other minor methodological flaws	Other important methodological flaws	

Appendix 2.1 – Reporting Checklist

	Recommendation (Streiner & Kotter, 2014)	Current Study
Title	· Clearly indicate whether the paper reports on the development of a new instrument or the testing of an existing one. · Specify the main focus (e.g., reliability, validity, or both).	The title explicitly states a newly developed measure of sibling relationship quality.
Abstract	· Provide a concise overview of aims, design, sample, methods, key results, and conclusions. · Avoid stating "the scale is valid" or "the scale is reliable" as absolute; rather suggest context or applications.	The abstract summarises aim, participant count, factor analysis methods, and key findings.
Keywords	· Include terms such as "instrument development," "reliability," "validity," or the construct being measured. · Consider relevant indexing terms (e.g., "scale validation," "sibling relationships," "psychometrics").	Keywords (e.g., "Sibling Relationship Quality," "Psychometrics," "Scale Development") are included.
Introduction	· Rationale for developing a new scale vs. using or adapting an existing one. · Conceptual/theoretical framework describing the construct (e.g., sibling relationship quality) · Gaps in existing literature that justify a new measure.	The introduction details the need for a new, comprehensive sibling measure. Constructs (Warmth, Conflict, Rivalry, etc.) are contextualized with theoretical and empirical justifications.
Methods	· Description of the Instrument · Number of items, subscales, scoring format, response options, any reverse-scored items. · Participants & Sampling: · Detailed inclusion/exclusion criteria, sampling strategy, description of participant characteristics (age, gender, setting). Justify sample size	Instrument: 187 initial items refined to 34, subscales identified Participants: Parents/guardians (N=457) with at least two children under 18; demographics detailed. Development Process: Three phases with expert consultations (Phase 1 & 2), large-scale data collection (Phase 3)

	(e.g., rules of thumb for factor analysis, reliability precision, or prior power calculations). · Instrument Development Process for new scales: item generation (focus groups, literature reviews, expert panels), content validation steps, pilot testing. · Procedure: Outline data collection timeframe, test–retest intervals if applicable, instructions given to participants, any blinding for inter-rater studies. · Ethical approval, informed consent, confidentiality measures, handling of incomplete data. · Statistical Analyses: List which reliability indices (e.g., Cronbach's alpha, ICC) and validity tests (e.g., factor analysis, convergent/discriminant evidence) were used. · Justify the factor extraction (e.g., EFA with ML extraction, parallel analysis). · Mention software and version used.	Procedure: EFA with ML extraction & Promax rotation; parallel analysis for factor retention. Ethics board details given (ID: 200230225) Analyses: SPSS v29, Cronbach's alpha, EFA loadings, correlations for validity.
Results	· Provide demographic info, flow diagram if relevant (response rates, exclusions) Reliability: Cronbach's alpha (with confidence intervals if possible), test–retest or inter-rater reliability, item-total correlations, or other relevant estimates. · Validity: Summaries of factor analyses (eigenvalues, loadings), correlations with other measures (convergent/discriminant), known-groups comparisons if done.	Sample: Demographics for Phase 2 (parents, clinicians, researchers) & Phase 3 (457 final sample). Reliability: Cronbach's alpha for each factor (range .82–.93); item reduction steps described. Validity: EFA loadings, factor correlation matrix (< .85), AVE for each subscale, parallel analysis justification. Detailed tables (e.g., Table 3, 4, 5, 6) showcasing factor loadings, alpha coefficients, descriptive stats, inter-factor correlations.

	· Present results in tables/figures for easy interpretation.	
Discussion	· Interpret the psychometric findings (e.g., factor structure, reliability levels), referencing prior research/theory. · Emphasize that reliability/validity are not fixed properties; they depend on sample and context. · Acknowledge limitations (e.g., representativeness, single-informant bias). · Propose future directions (e.g., confirmatory factor analysis, multi-informant designs, cultural validations).	Discusses factor structure aligning with established constructs + new dimensions (e.g., Mentalising). Notes reliability & validity vary by population/setting. Limitations: EFA only; skewed towards White, educated sample; reliance on parent reports. Future directions: CFA, multi-informant approaches, diverse samples, longitudinal data.
Conclusion	· Avoid absolute statements like "the scale is valid." · Summarize key outcomes (e.g., potential for clinical/research use, suggestions for ongoing validation).	Concludes that SRQ-IR is promising with distinct subscales and strong psychometrics, but further research (CFA, multi-informant) is needed.
Practical/Clinical Implications	· State how the instrument can be used in practice or research (e.g., identifying sibling dynamics, designing interventions). · Note that improved reliability/validity alone does not guarantee better outcomes unless tested in real-world settings.	Highlights how SRQ-IR can help clinicians and researchers identify sibling relational strengths and challenges; suggests it may guide interventions.
Limitations & Future Research	· Provide transparent discussion of sample constraints (e.g., size, demographics), potential biases, measurement boundaries. · Outline next steps for validation in broader or different contexts.	Acknowledges potential issues (e.g., partial homogeneity of sample, single-informant design), encourages replication with diverse groups.
References	· Follow a consistent, recognised citation style (e.g., APA, Vancouver). · Include references for key psychometric guidelines (e.g.,	Key references are included APA style

	Clark & Watson, relevant scale development texts).	

Appendix 2.2 – Final Approved MRP Proposal

Final Approved MRP Proposal can be accessed at the following link:

https://osf.io/2mvjt/files/osfstorage/67b754217529e7c171c34d35

Appendix 2.3 – Project Approval Letter

University of Glasgow

Professor Hamish McLeod

MVLS College Ethics Committee

Developing and validating a new scale of sibling relationship quality *200230225*

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study.

We are happy therefore to approve the project, subject to the following conditions

- Project end date as stipulated in original application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or groups or datasets as defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- For projects requiring the use of an online questionnaire, the University has an Online Surveys account for research. To request access, see the University's application procedure at https://www.gla.ac.uk/research/strategy/ourpolicies/useofonlinesurveystoolforresearch/.
- You should submit a short end of study report within 3 months of completion.

Yours sincerely

Dr Terry Quinn

Terry Quinn
FWSO, FESO, MD, FRCP, BSc (hons), MBChB (hons)

College of Medicine, Veterinary & Life Sciences
School of Cardiovascular and Metabolic Health
New Lister Building, Glasgow Royal Infirmary
Glasgow G31 2ER
terry.quinn@glasgow.gla.ac.uk
Tel – 0141 201 8519

The University of Glasgow, charity number SC004401

Appendix 2.4 – Participant Information Sheets and Consent Form

Participant information sheets and consent form can be accessed at the following links:

Phase 2 Parents, Clinicians and Researchers Information sheets:

https://osf.io/2mvjt/files/osfstorage/67b75599c8f79818f90c77ba

https://osf.io/2mvjt/files/osfstorage/67b755fec8f79818f90c77f6

https://osf.io/2mvjt/files/osfstorage/67b75610a8b3a521aa3137f1

Consent Form:

https://osf.io/2mvjt/files/osfstorage/67b75568c8f79818f90c779c

Appendix 2.5 – Sibling Relationship Quality – Informant Report Scale

Sibling Relationship Quality – Informant Report

Below are statements regarding the relationship between two children who are siblings. This questionnaire is designed for completion by an adult who is familiar with the sibling pair. Please respond based on your observations and knowledge of their interactions. Your input will help us understand the dynamics and strengths of their sibling relationship.

Please carefully read each statement and assess how often each one accurately describes the sibling relationship **over the last month.**

1	2	3	4	5
Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**

Over the last month…							
1. They like each other	1	2	3	4	5		WC
2. They deliberately annoy each other	1	2	3	4	5		A
3. They grass each other up (tell on each other)	1	2	3	4	5		A
4. Difficult times bring them closer together	1	2	3	4	5		WC
5. They try to think about how each other feels	1	2	3	4	5		M
6. One thinks the other is more talented (child one or two?)	1	2	3	4	5		R
7. They act as if they are on the same team	1	2	3	4	5		WC
8. They are able to find compromises together	1	2	3	4	5		CR
9. They try to think about how each other feels	1	2	3	4	5		M
10. They tease each other	1	2	3	4	5		A
11. They seek help from adults or older children when conflicts become challenging to resolve	1	2	3	4	5		CR
12. They take care of each other	1	2	3	4	5		WC
13. They argue with each other	1	2	3	4	5		A

14. One attempts to sabotage the other (child one or two?)	1	2	3	4	5		RA
15. They know why each other acts the way they do	1	2	3	4	5		M
16. They do not need a lot of prompting to apologise to each other after an argument	1	2	3	4	5		CR
17. One can be unnecessarily cruel to the other (child one or two?)	1	2	3	4	5		RA
18. One thinks the other is better (child one or two?)	1	2	3	4	5		R
19. They try to think about the reasons why each other behaves the way they do	1	2	3	4	5		M
20. They choose to spend their free time together	1	2	3	4	5		WC
21. One makes the other say or do things they do not want to do (child one or two)	1	2	3	4	5		RA
22. They deliberately annoy each other	1	2	3	4	5		A
23. They are physically affectionate with each other	1	2	3	4	5		WC
24. They know why each other acts the way they do	1	2	3	4	5		M
25. One causes fear and alarm towards the other (child one or two?)	1	2	3	4	5		RA
26. One feels overshadowed by the other (child one or two?)	1	2	3	4	5		R
27. They are interested in guessing what the other is feeling	1	2	3	4	5		M
28. One displays developmentally inappropriate sexual behaviours towards the other (child one or two?)	1	2	3	4	5		RA
29. They try and listen to each other during a conflict	1	2	3	4	5		CR
30. They get on each other's nerves	1	2	3	4	5		A
31. One thinks the other is more successful (child one or two?	1	2	3	4	5		R
32. One attempts to turn other children against the other (child one or two?)	1	2	3	4	5		RA
33. They are able to be friendly to each other after an argument has been resolved	1	2	3	4	5		CF
34. They try to see situations through the eyes of the other	1	2	3	4	5		M

Total Scores:

Antagonism: ____
Warmth/Closeness: ____
Relational Aggression: ____

Conflict Resolution: ____
Mentalising: ____
Rivalry: ____

Demographic Information

Child 1
Age: _____
Gender: ☐ Male ☐ Female ☐ Non-binary ☐ Other _____

Child 2
Age: _____
Gender: ☐ Male ☐ Female ☐ Non-binary ☐ Other _____

Relationship between Child 1 and Child 2:
☐ Full siblings ☐ Half-siblings ☐ Step-siblings ☐ Adoptive siblings ☐ Other _____

Your relationship to these children:
☐ Parent ☐ Guardian ☐ Teacher ☐ Clinician ☐ Other _____

Appendix 2.6 – Data Analysis Plan

The data analysis plan can be accessed at the following link:

https://osf.io/2mvjt/files/osfstorage/67b757ddfce933b17da54d1b

Appendix 2.7 – Records of Data Analysis Process

Records of the data analysis process and data decisions and rationales can be accessed at the following links:

Data analysis process (SPSS Syntax):

https://osf.io/2mvjt/files/osfstorage/67b773271605653208c34fa1

Data decisions and rationales:

https://osf.io/2mvjt/files/osfstorage/67b758af2c3691b9f8c34ef5

Appendix 2.8 – Data Availability Statement

The data supporting this study's findings are available on request from the lead researcher. The data are not publicly available due to privacy or ethical restrictions.