ORIGINAL PAPER

CrossMark

How Children and Adolescents Evaluate their Families: Psychometric Properties and Factor Structure of the Evaluation of Social Systems (EVOS) Scale

Dennis Grevenstein[1] · Jochen Schweitzer[2] · Corina Aguilar-Raab [2]

Abstract

The Evaluation of Social Systems (EVOS) scale was developed as a measure of the quality of social relationships vis-à-vis context-unspecific aspects of relationship quality. Although constructed for adolescents above the age of 11, it has to date only been validated in adult samples. The current study investigated if EVOS can be applied to adolescents through a cross-sectional study. We investigated factorial validity and construct validity of EVOS in a sample of adolescents with a mean age of 14.74. Additionally, we examined measurement invariance between youth and adult samples. The two-factor model that was developed for adults also held in the present study. EVOS showed good reliability and exhibited strict measurement invariance. Construct validity was supported by correlations with measures of family functioning and psychological distress. We conclude that EVOS is an economic, reliable, and valid measure of the quality of family relations in adolescents. Scores are directly comparable between youths and adults.

Keywords Quality of relationships · Family therapy and counseling · Measurement invariance · Construct validity · reliability

Family therapy has been shown to be similarly effective compared to individual therapy, exhibiting moderate to high overall effect sizes (Heatherington et al. 2015; Sexton et al. 2013). In a systematic review of 47 randomized controlled trials evaluating systemic therapy for externalizing disorders in childhood, including attention deficit hyperactivity disorder and substance use disorders, 42 studies have shown the efficacy of systemic therapy for specific symptom reduction (von Sydow et al. 2013). Another 33 out of 38 randomized controlled trials of systemic family therapy supported its efficacy for internalizing and mixed disorders in childhood and adolescence, such as mood disorders (Retzlaff et al. 2013). Outcome research indicates that systemic therapy is almost equally efficacious compared to well-established family therapy approaches such as Cognitive-Behavioral Therapy and Psychodynamic Therapy (Carr 2014b; Pinquart et al. 2016).

Family therapy targets social and interactional processes between all members. Yet, it is worth noting that emotional or behavioral problems of children are one of the most common reasons for families to seek treatment (Merikangas et al. 2010; Merikangas et al. 2009). In many cases the child or adolescent is considered to be the *index*-patient of the family system (i.e. the patient carrying the symptom). The main focus in outcome research is usually the symptom itself. Here, observational measures using for example coding schemes to analyze play tasks or specific parent-child interactions in laboratory or other settings can be used to assess treatment outcome in detail (Aspland and Gardner 2003). Questionnaires are more economic and a number of measures are frequently used to evaluate improvements in the child's or adolescent's symptom load. The Child Behavior Checklist (Achenbach 1991; Achenbach and Rescorla 2004) focuses on psychosocial (adaptive)

✉ Corina Aguilar-Raab
corina.aguilar-raab@med.uni-heidelberg.de

[1] Psychological Institute, University of Heidelberg, Heidelberg, Germany

[2] Institute of Medical Psychology, Center for Psychosocial Medicine, University Hospital, Heidelberg, Germany

functioning/dys-functioning, problems, and competencies. The Behavior Assessment System for Children-2 (Reynolds and Kamphaus 2004) is directed towards emotional or behavioral disorders. Finally, the Youth Outcome Questionnaire 2.01 (Atkin et al. 2001; Burlingame et al. 2005; Burlingame et al. 2004) parallels youth self-report version (Wells et al. 2003) and assesses behavior change resulting from therapy (McClendon et al. 2011).

By referring to systemic theory, the focus on the symptom presents a problem. Even if the child or adolescent is the *index*-patient of the family system, systemic approaches consider the interactions and relationships between all system members as crucial factors in the development and sustainment of a disease. Unfortunately, only cautious attempts have been made to include interactional aspects as outcome variables in psychotherapy research in general. Yet, focusing only on symptom reduction can mislead especially when an intervention targets other aspects such as social interactions or communication styles (Adelman et al. 2014). Social interactions, social support, relationship quality and relationship satisfaction are important predictors for health and well-being (Miller et al. 2009; Rusbult and Van Lange 2003). Relationship quality and relationship satisfaction between children/adolescents and their parents are not just important for child/adolescent development (Cleveland et al. 2008; Moore et al. 2004) but can also foster family and individual resilience (Patterson 2002; Rayner and Montague 2000; Walsh 2003). The parent-child relationship changes along the lifespan as reliance on the parents diminishes and changes into a more reciprocal relationship (Allen et al. 2003; Noller et al. 2001).

Different theories aim to describe and explain important domains of relationship quality in families, referring to well-being and social-functioning and taking into account the changing needs of the developing child and adolescent. The sense to belong (Resnick et al. 1993) functions as a protective factor for different problem and high risk behaviors. A high sense of social connectedness among people but also to one's own family is strongly associated with higher well-being in adolescence (Jose et al. 2012). Low family cohesion is connected to depressive feelings, especially feelings of loneliness, and reduced social acceptance (Johnson et al. 2001). Family cohesion also predicted children's internalizing and attention problems longitudinally (Lucia and Breslau 2006). It follows that positive family involvement and expressed warmth are essentials for the establishment of a good family atmosphere. Both have been identified as predictors for reduced symptom development and better social functioning during the prodromal phase of schizophrenia (O'Brien et al. 2006). Perceived (social) support and the ability to get responsibly involved in family decisions provide a buffering against negative psychological outcomes or maladjustment (Cook et al. 2002; Demaray and Malecki 2002).

The frequency of communicational interaction along with the need for shared experiences and emotional self-disclosure diminish during the transition from childhood to adolescence. Still, the value of the continuity of (caring) parent communication for health outcomes has been frequently reported (Ackard et al. 2006; Laursen and Colling 2004; Phillips-Salimi et al. 2014). Collective efficacy is the perceived ability of the parent-child group to jointly and effectively perform a certain task using their skills based on a sense of cohesion (Bandura et al. 2011; Wells et al. 2004). This includes the ability for conflict management, problem solving and shared decision making fostering motivation, endurance, and overall family functioning. For example, collective efficacy enhances family attachment and support, protecting adolescents from suicidal behavior (Brooks-Gunn et al. 2010). It also includes the foundational ability to adapt to changing life experiences in different domains. Taken together, family functioning and the relationship quality between children/adolescents and their parents exhibit numerous health benefits. Adolescents perceive a positive, respectful and caring relationship with their parents as meaningful and helpful. This strongly encourages the inclusion of relational aspects in psychotherapy outcome research.

However, currently available measures of family functioning still have shortcomings. Based on a current systematic review, only five out of eight existing commonly used measures are suitable for assessing clinical outcome in couple and family therapy (Carr and Hamilton 2016): The McMaster Family Assessment Device (Baldwin et al. 1983; Mansfield et al. 2015), Family Adaptability and Cohesion Evaluation Scales (Olson 1985, 2011), Self-Report Family Inventory (Beavers and Hampson 2000), Family Assessment Measure III (Skinner et al. 1983; Skinner et al. 2000), and the Systemic Clinical Outcome Routine Evaluation (Carr and Stratton 2017; Stratton et al. 2010). Yet, they contain a large number of items, take a long time to complete, focus on a specific theory, are problem-oriented, and are usually context-specific. Moreover, only a few report their applicability to children and adolescents. Reports on quality criteria, including indices for sensitivity of change, are scarce (Olson et al. 1979).

The Evaluation of Social Systems scale (EVOS) was developed precisely to address the need for a brief and economic measure of the quality of social relationships for the evaluation of systemic, multi-person intervention outcomes in different social systems (such as families) from the perspective of all members of a targeted social system (Aguilar-Raab et al. 2018; Aguilar-Raab et al. 2015). EVOS aims to assess the valuable perspective of each system member without the presumption of "healthy" or "functional" relationships. The item wording theoretically originated from systems theory and from models of functionality

and relationships in families, such as the Beavers Systems Model (Beavers 1981), the Circumplex Model of Marital and Family Systems (Olson 1985; Olson et al. 1979), and the McMaster Model of Family Functioning (Baldwin et al. 1983; Bishop et al. 1978). Additionally, groups and social contexts with regard to organizational psychology (Anderson and West 1996; Faulstich 1998; Kauffeld 2001; McGrath et al. 2000), as well as Bandura's concept of collective efficacy were taken into account (Bandura et al. 2011). Overlapping dimensions were extracted in order to condense the most important ones for different social systems. In line with systems theory, EVOS follows a non-problem-oriented, non-normative, and constructivist approach. It focuses on the first person plural "we…" instead of "I…", in order to assess an individual's perception of the group or system as a whole, including the individual him/herself. The EVOS scale includes nine items covering the following aspects subsumed under the two factors (1) *quality of relationship* and (2) *collective efficacy* —all of which are important aspects for the quality of child/adolescent-parent relationship: (1) communication, cohesion, atmosphere, giving and taking, as well as (2) collective aims, resources, decision making, finding solutions, and adaptability. An additional, separate tenth item, that is not used to calculate the EVOS total score, can assess the perceived consensus within the system for additional evaluation. The EVOS scale was initially developed using exploratory factor analysis resulting in two distinct factors. On the basis of systems theory, we assumed the two factors (i.e. quality of relationship and collective efficacy) to be correlated, rather than being independent from each other. Additionally, EVOS aimed at a global evaluation of the quality of a social relationship. These assumptions were tested using confirmatory factor analysis. A model with two correlated factors was established in adult samples and validated in both English and German. EVOS has also been evaluated in different social contexts—such as couples, families and working teams—and proved to be equally applicable to all social contexts. Thus, the scale is a short, reliable, sensitive to change, and valid measure of the quality of relationship and collective efficacy (Aguilar-Raab et al. 2018; Aguilar-Raab et al. 2015).

In order to monitor family therapy sessions or evaluate the outcome of family therapy, all members of the family— including adolescents— should be assessed. We believe that children and adolescents are especially important as pillars of a family life. Disregarding them in family therapy outcome research means to miss an integral part of a picture— in such cases, findings can only be misleading. In order to respect children and adolescents, and their associated needs for a supportive environment, it is vital to apply measures that reflect and include their views. Commonly used self-rating measures of family functioning have rarely been tested for applicability in younger ages. Additionally, it is most often unclear if ratings provided by children and adolescents are truly comparable to adults' judgments. The present research investigated the applicability of EVOS in a sample of adolescents. We examined factorial validity, construct validity, and measurement invariance between adolescent and adult samples.

Method

Participants

Participants were $N = 203$ adolescents from two different schools in Germany. In Germany, there are three different tiers of schools. Basic schooling ("Hauptschule") goes up to the ninth grade. Middle school ("Realschule") means schooling up to the tenth grade and finally, high school ("Gymnasium") is the university-track school type and means schooling up to the 13th grade. The first subsample encompassed $n = 112$ middle school students ($M_{age} = 14.51$; $SD = 1.62$; range = 12 to 18; 46.4% female). The second subsample constituted $n = 91$ high school students ($M_{age} = 15.02$; $SD = 2.18$; range = 11 to 19; 48.4% female).

Procedure

For practical reasons and ethical concerns on the side of school administration, participants in subsample 1 completed only the EVOS. Adolescents in subsample 2 filled out additional validation measures. Participants, parents, and teachers were fully informed about the project. Prior to participation, adolescents and parents provided informed consent. Participants could individually take part in a lottery for vouchers as an incentive. Additionally, the school class as a whole could participate in a lottery for a lunch order of their choice for all students in class. For comparison purposes, we used data presented by Aguilar-Raab et al. (2015), where $N = 188$ adults with a mean age of 30.46 ($SD = 13.07$; 80.9% female) evaluated their family relationship. The study was approved by the ethics committee of the university hospital Heidelberg (S-508/2012).

Measures

EVOS: Family relationships

Family relationship quality was assessed using EVOS (Aguilar-Raab et al. 2015). The quality of the social system is assessed by four items capturing the emotional/affective level of the quality of the relationship (e.g., "For me, the way we talk with each other, is …"), and five items measuring the rather cognitive evaluation of collective efficacy

of the family (e.g., "For me, how we adapt to change, is …"). Answers on 4-point rating scales range from 0 = *very poor* to 3 = *very good*. We computed three different scores: We computed mean scores for both subscales quality of the relationship (Cronbach's Alpha = 0.81) and collective efficacy (Alpha = 0.80) as well as a mean score for the whole scale (Alpha = 0.88). EVOS includes a tenth item measuring the perceived consensus among family members. This item is not an integral part of the scale and may be used to explicitly assess the agreement between family members. For the purpose of the present study, where we have only a single assessment of a certain family, we ignored the consensus item.

SCORE-15: Family functioning

SCORE-15 is a validated measure of family functioning assessing strengths, difficulties, and communication among family members (Hamilton et al. 2015; Stratton et al. 2014). Sample items include "People often don't tell each other the truth in my family" and "It feels miserable in our family". Answers were given on 5-point rating scales marked from 0 = *do not agree at all* to 4 = *completely agree*. A mean score was computed. Cronbach's Alpha amounted to .81.

SDQ-deu: Psychological distress

We used a German adaptation (Lohbeck et al. 2015) of Goodman's (1997) Strengths and Difficulties Questionnaire (SDQ). The scale is a validated instrument with 25 items, measuring social behavior, emotional problems, behavioral problems, hyperactivity, and problems with peers. Sample items include "I worry a lot" and "I often get into fights. I can force others to do what I want". Each of the five subscales contains five items. Answers were given on three-point scales marked 0 = *not true*, 1 = *partially true*, 2 = *completely true*. Negatively coded items were reversed before summing up items to subscale and total scale scores. Higher scores indicate greater difficulties. Scale reliabilities amount to Alpha = .69 for the SDQ total score, as well as .73 (emotional problems), .38 (behavioral problems), .71 (hyperactivity), .70 (peer problems), and .63 (social behavior) for the subscales.

Data Analysis

We used SPSS 22 (IBM, 2012) and Mplus 7.11 (Muthén and Muthén 1998-2012). For confirmatory factor analyses model fit was evaluated by (1) the—ideally non-significant — χ^2 test (Bentler and Bonett 1980); (2) the comparative fit index (CFI) with values of .90/.95 and above indicating appropriate/good model fit (Bentler 1990; Hu and Bentler 1999); (3) the root mean square error of approximation

(RMSEA) with values less than 0.08 indicating good model fit (Browne and Cudeck 1993); and (4) the standardized root mean square residual (SRMR) with values less than 0.08 (Hu and Bentler 1999) considered to reflect good fit. For model comparisons we used the Bayesian Information Criterion (BIC) as a comparative fit index (Schwarz 1978). Lower scores indicate better model fit and differences greater than +/-10 indicate unequal fit (Raftery 1995).

We initially checked for multivariate normality using Small's omnibus test. Results indicated that multivariate normality did not hold, both in the adult (χ^2 (18) = 78.54, $p < 0.001$) as well as in the adolescent sample (χ^2 (18) = 154.76, $p < 0.001$). This was expected due to EVOS' coarse 4-point scale and is in line with prior research on the EVOS scale (Aguilar-Raab et al. 2015). Hence, we used a full information maximum likelihood estimator with robust standard errors (MLR) in all CFAs. Mplus provides MLR for maximum likelihood with robust 'Huber-White' standard errors and a scaled test statistic asymptotically equivalent to the Yuan–Bentler T2* statistic (Yuan and Bentler 2000) and similar to the robust Satorra–Bentler scaled χ^2-statistic (MLM; Chou et al. 1991). A total of 0.18% of EVOS data were missing at random and subsequently handled using full information ML. The validation measures had 0.83% missing values. Due to the negligible amount of missing data, we computed mean scores for all participants and variables. A subsample of only $n = 90$ was available for correlation analyses. Nonetheless the sample size should be adequate to investigate construct validity. In adult samples, quality of family and couple relationships (as measured with EVOS) correlated with measures of general psychopathology between $r = -0.30$ up to $r = -0.54$ (Aguilar-Raab et al. 2018; Grevenstein et al. 2018). We expected to find a similar correlation in adolescent samples. Power calculation using G-Power (Faul et al. 2007) indicated that a sample size of $N = 84$ would be sufficient to detect a correlation of $r = -0.30$ assuming common parameters of α = 0.05 (two-tailed) and power $(1 - β) = 0.80$. For single-sided testing, which is clearly in line with our hypothesis, the required sample size is even reduced to $N = 67$.

To compare the measurement model of EVOS between adolescent and adult samples, we examined measurement invariance (MI) (Vandenberg and Lance 2000). When applying a scale across situational contexts, cultures, or age groups with the intent to compare their scores numerically, most researchers simply assume that the scores reflect the identical construct. Yet, differential use of a scale and disparate measurement models in specific samples may also account for differences between groups. Mean scores or correlations with external measures can only be meaningfully compared if MI can be established (Chen 2008). One needs to ascertain that differences in scale means are

Table 1 Correlations between study variables

	EVOS tot	EVOS RQ	EVOS CE	SCORE-15	SDQ tot	SDQ EP	SDQ BP	SDQ HYP	SDQ PP	SDQ SB
EVOS total	**0.88**									
EVOS relationship quality	0.90[**]	**0.81**								
EVOS collective efficacy	0.94[**]	0.71[**]	**0.80**							
Family functioning (SCORE-15)	0.78[**]	0.74[**]	0.69[**]	**0.81**						
SDQ total score	−0.33[**]	−0.35[**]	−0.26[*]	−0.45[**]	**0.69**					
SDQ emotional problems	−0.29[**]	−0.34[**]	−0.20	−0.45[**]	0.70[**]	**0.73**				
SDQ behavioral problems	−0.25[*]	−0.26[*]	−0.21[*]	−0.35[**]	0.64[**]	0.20	**0.38**			
SDQ hyperactivity	−0.06	−0.06	−0.05	−0.10	0.59[**]	0.12	0.31[**]	**0.71**		
SDQ peer problems	−0.25[*]	−0.24[*]	−0.22[*]	−0.24[*]	0.64[**]	0.29[**]	0.39[**]	0.04	**0.70**	
SDQ social behavior	0.19	0.20	0.15	0.07	−0.17	0.06	−0.24[*]	−0.10	−0.23[*]	**0.63**

Note: Scale reliabilities (Cronbach's Alpha in bold print) are depicted in the diagonal

[*] $p < 0.05$, [**] $p < .01$.

due to true differences in latent means, not different item utilization.

MI is tested within a sequential approach of nested, increasingly restricted confirmatory factor-analytical (CFA) models. Four increasingly restrictive forms of MI are usually tested (Meredith 1993; Schmitt and Kuljanin 2008; Vandenberg and Lance 2000) (1) Configural MI assumes equal construct dimensionality and equivalent item-to-factor patterns across groups. (2) Under Metric MI, all item loadings are constrained to be equal across groups, indicating that the same construct is measured in both groups and that participants attribute the same meaning to the latent construct. (3) Scalar MI additionally assumes invariant item intercepts across groups, indicating that scores are based on the same unit of measurement. Scalar MI allows a meaningful interpretation of latent mean differences. (4) Strict MI additionally requires equality of item residuals, indicating equal reliability across groups. When strict MI holds, all differences on manifest variables are due to true differences on the latent variables, rather than measurement error. Strict MI would then allow a direct comparison of observed mean scores across groups. From this level onward, one can additionally examine the invariance of structural parameters, i.e. latent means and factor variances. If some parameters are non-invariant across groups, a weaker form of MI, partial invariance, may still hold. For example, partial scalar MI requires most, but not all, of the item intercepts to be invariant across groups. In this case latent means may still be cautiously compared (Byrne et al. 1989; Lubke and Dolan 2003).

When testing for MI, models are compared using χ^2-difference tests. Due to the use of robust estimation procedures (MLR), we used Satorra-Bentler scaled χ^2-difference tests (Satorra 2000; Satorra and Bentler 2001). χ^2-tests are often highly dependent on sample size, so fit indices are commonly used to judge MI. Differences of ΔCFI and ΔRMSEA are examined and a drop in CFI less or equal to .010 is considered acceptable, as long as it is balanced by ΔRMSEA no greater than +0.015 (Chen 2007; Cheung and Rensvold 2002). Finally, lower BIC values indicate a better tradeoff between accuracy and parsimony.

Results

We initially examined if both adolescent subsamples differed. There were no significant differences with regard to participant sex, $\chi^2 = 0.07$, $df = 1$, $p = 0.89$, or age, $t = 1.86$, $df = 160.48$, $p = 0.065$, $M_1 = 14.51$, $SD_1 = 1.62$, $M_2 = 15.02$, $SD_2 = 2.18$, $d = 0.27$. Still, sample 1 reported a lower EVOS mean score than sample 2, $t = 2.20$, $df = 201$, $p = 0.029$, $M_1 = 2.29$, $SD_1 = 0.50$ vs. $M_2 = 2.45$, $SD_2 = 0.51$, $d = 0.32$. Correlations between study variables can be seen in Table 1. Means and standard deviations are depicted in Table 2. Overall our sample was comparable to other adolescent samples. The SDQ scores closely matched the scores reported for a recent German standardization (Lohbeck et al. 2015).

A total of $N = 200$ participants could be used for psychometric analyses due to missing values (listwise-deletion is required). The items tended to be answered (more) positively, as would be expected in general school samples (all Ps > 74.42). Corrected item-to-total correlations ranged between 0.52 and 0.72, indicating good reliability of the EVOS. Cronbach's Alpha for the whole scale amounted to 0.88 for the whole scale and 0.80 for the subscale relationship quality and 0.81 for the subscale collective efficacy. This closely resembled scale reliabilities in adult

Table 2 Descriptives of study variables

	n	M	SD	Range
Family relationship quality (EVOS) adults	188	1.93	0.66	0–3
EVOS relationship quality (adults)	188	2.09	0.71	0–3
EVOS collective efficacy (adults)	188	1.80	0.70	0–3
Family relationship quality (EVOS) adolescents	203	2.38	0.51	0.56–3.00
EVOS relationship quality (adolescents)	203	2.45	0.54	0.50–3.00
EVOS collective efficacy (adolescents)	203	2.32	0.56	0.60–3.00
Family functioning (SCORE-15)	91	3.13	0.49	1.67–4.00
Psychological distress (SDQ-deu total score)	90	10.78	5.13	3–30
Emotional problems (SDQ-deu)	90	3.12	2.50	0–10
Conduct problems (SDQ-deu)	90	1.87	1.39	0–7
Hyperactivity (SDQ-deu)	90	3.60	2.20	0–10
Peer problems (SDQ-deu)	90	2.19	1.89	0–7
Social behavior (SDQ-deu)	90	8.16	1.69	0–10

Table 3 EVOS item descriptives and factor loadings

Items	Theoretical dimension	M	SD	P	r_{itc}	λ CFA
1	Communication	2.49	0.69	83.25	0.66	0.74
2	Cohesion	2.48	0.69	82.76	0.64	0.74
3	Giving and taking	2.47	0.65	82.43	0.62	0.68
4	Atmosphere	2.36	0.70	78.65	0.62	0.70
5	Aims	2.24	0.75	74.42	0.61	0.64
6	Resources	2.27	0.78	75.62	0.62	0.69
7	Decisions	2.39	0.73	79.47	0.56	0.64
8	Extending perspectives	2.39	0.76	79.80	0.74	0.83
9	Adaptability	2.32	0.76	77.34	0.52	0.56

Note: $N = 200$

samples, where Alphas of 0.87 for the total score and 0.82 for the subscale relationship quality and 0.82 for the subscale collective efficacy have been reported.

To test the measurement model of the EVOS in adolescent samples, we used confirmatory factor analysis. EVOS was developed as a scale implementing a construct with two theoretically distinct, yet highly related factors using exploratory factor analysis. Consequently, a model with two correlated factors for quality of the relationship and collective efficacy was established for adult samples. In the adult sample, this two-factor model was shown to fit the data (Aguilar-Raab et al. 2015), $\chi^2 = 38.10$, $df = 26$, $p = 0.06$; RMSEA $= 0.05$ CI$_{90}$ = [0.00–0.08], p-close $= 0.47$; CFI $= 0.985$; SRMR $= 0.030$; AIC $= 3344$; BIC 3434. We now aimed to replicate the previous findings with our new data. The two-factor mode also fitted the adolescent data very well, $\chi^2 = 25.37$, $df = 26$, $p = 0.50$; RMSEA $= 0.00$ CI$_{90}$ = [0.00–0.05], p-close $= 0.93$; CFI $= 1.00$; SRMR $= 0.031$; AIC $= 3338$; BIC 3431. An attentive reader will notice that this model achieved perfect model fit in terms of CFI and RMSEA due to χ^2 being less than the degrees of freedom. This is a result of using a robust maximum likelihood estimator, which produces scaled χ^2 test statistics. If χ^2 is less than df, which is the expected χ^2 in the population, then RMSEA is set at zero, indicating a perfectly specified model. Similarly, if CFI is greater than one, it is set at one and if less than zero, it is set at zero (see Kenny 2015). We computed the average variance extracted (AVE; Fornell and Larcker 1981) for the two sub-factors (quality of the relationship: AVE $= 0.51$; collective efficacy: AVE $= 0.46$) as

well as for the whole scale, AVE $= 0.48$. Both factors were highly correlated at $r = 0.87$, which may be construed as an indicator of poor discriminant validity. Thus, we also checked the applicability of a single factor model. This model also fitted the data, $\chi^2 = 54.04$, $df = 27$, $p < .002$; RMSEA $= 0.07$ CI$_{90}$ = [0.04–0.10], p-close $= 0.11$; CFI $= 0.961$; SRMR $= 0.038$; AIC $= 3357$; BIC 3446. Even though a single factor model also fitted the data, the two-factor model was superior as was clearly visible in all the indicators of model fit, both absolute and comparative. The item descriptives and factor loadings of the two-factor model of EVOS are depicted in Table 3 and the structural model with standardized loadings can be seen in Fig. 1.

The initial step in MI testing indicated that configural MI held across age groups. Metric MI also held unconditionally. When testing for scalar MI, model fit noticeably decreased. We then examined if the invariance was caused by specific parameters. Relaxing one candidate parameter at a time on the basis of modification indices (Byrne et al. 1989), we checked if model fit could be improved. The fit of the scalar MI model improved noticeably after relaxing equality constraints for the intercept of item #1 ("For me, the way we talk with each other, is …" (ModInd = 17.12). Apparently, this item was easier to endorse for adolescents. Thus, partial scalar invariance could still be shown. At the next step, we tested for equal item residuals. No relevant decrease in model fit could be observed, thus strict MI held across samples. Hence, EVOS showed comparable reliability for adolescents and adults.

Based on these positive results, we examined the invariance of structural properties. First, we constrained both factor variances and the covariance between factors to be equal across age groups. Model fit was slightly diminished, even though it might still be within conventionally accepted limits. Yet SRMR clearly indicated a non-fitting model, so we carefully refrain from declaring equal (co)variances. In the last step we additionally

Fig. 1 Measurement model of EVOS with standardized factor loadings
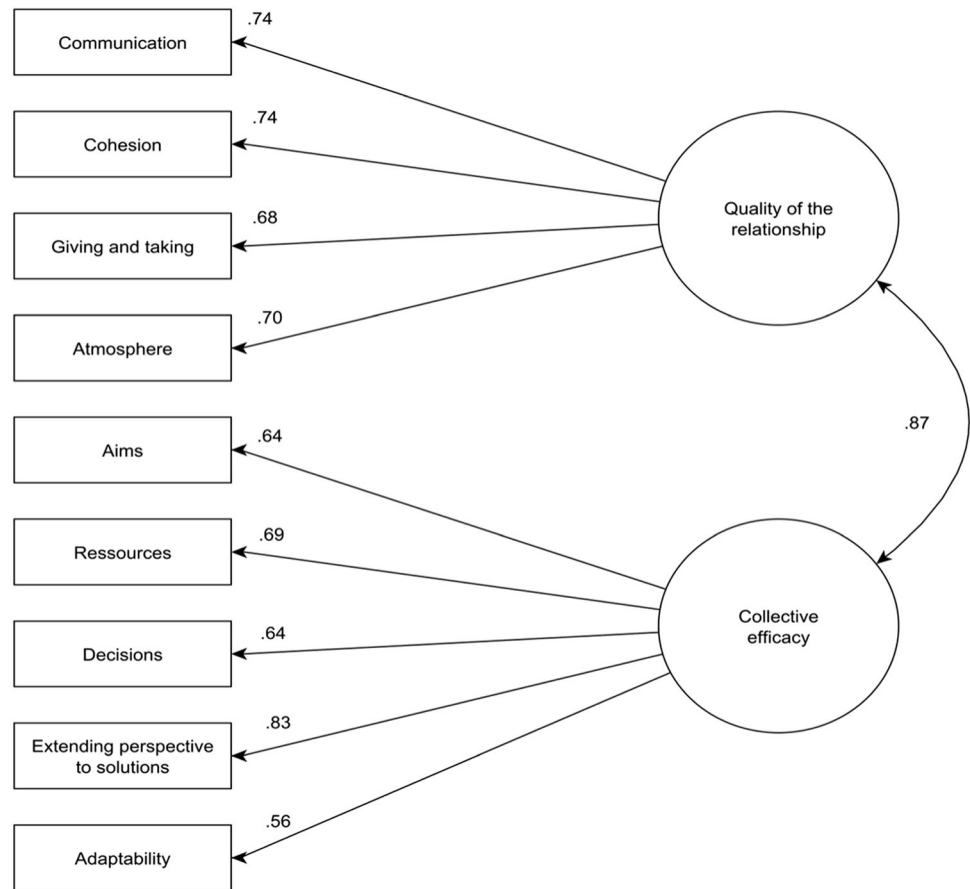
constrained latent means to be equal. Once again, model fit dropped substantially, because means were not equal. This was mirrored by a comparison of observed mean scores between adolescents and adults ($d = 0.76$). The measurement invariance for EVOS across adolescent and adult samples according to multiple group confirmatory factor analysis can be seen in Table 4.

We examined correlations between EVOS and the other constructs to assess construct validity in the second subsample. Both subscales of the EVOS showed similar correlations to the validations measures, thus we tend to interpret the global score, rather than the subscales. EVOS and SCORE-15 correlated at $r = 0.78$, indicating a strong overlap and convergent validity. The correlation between EVOS and SDQ was noticeably weaker ($r = -0.33$). Still, better perceived family relations indicated less psychological symptoms. Similarly, SCORE-15 correlated with psychological distress at $r = -0.44$. Correlations with psychological distress did not significantly differ for EVOS and SCORE-15, $Z = 1.70$, $p = 0.09$, even when accounting for the variance overlap between EVOS and SCORE-15 (Lee and Preacher 2013; Steiger 1980).

Discussion

Family therapy gives room for all involved members to express their views, needs, and feelings with regard to their relationship, the functioning of the family system as a whole and one's individual experience within this particular social system. Nonetheless, studies on family therapy research considering youth perspective are scarce (Moore and Seu 2011). Overall, children and adolescents are relegated within the mental health system. As a result, most research disregards youths' feedback, even though perspectives of parents and their children differ considerably (Celinska et al. 2015). Obviously, children and adolescents influence their social contexts and relationships as active agents and create their own perspectives based on individual experiences. To meet criteria for participatory research, it is necessary and beneficial to include all involved members of a focused (social) system. Doing so not only increases validity but also enhances the fit between mental health services and those who make use of it (Calheiros and Patrício 2014). The involvement of family members in treating disorders in young people is highly recommended

Table 4 Measurement invariance for EVOS across adolescent and adult samples according to multiple group confirmatory factor analysis (MGCFA)

MGCFA comparison invariance level	Equal loadings	Equal intercepts	Equal residuals	Equal variances	Equal means	df	χ^2	Δdf	$\Delta \chi^2$	p	CFI	RMSEA [CI_{90}]	SRMR	BIC
1 Configural						52	62.90	–	–	–	0.991	0.033 [0.000–0.059]	0.031	6682
2 Metric	×					59	70.37	7	7.43	0.386	0.991	0.031 [0.000–0.056]	0.048	6872
3 Scalar	×	×				66	98.08**	7	30.90	<0.001	0.975	0.050 [0.027–0.070]	0.054	6863
3a Partial scalar	×	×				65	80.39	6	10.49	0.105	0.988	0.035 [0.000–0.058]	0.054	6847
4 Strict	**×**	**×**	**×**			**74**	**95.35**	**9**	**15.39**	**0.081**	**0.983**	**0.038 [0.004–0.059]**	**0.055**	**6810**
5 Equal variances	×	×	×	×		77	110.16**	3	13.64	<0.001	0.974	0.047 [.025–.066]	0.144	6812
6 Full structural	×	×	×	×	×	79	164.98***	2	65.81	<0.001	0.932	0.075 [0.059–0.091]	0.245	6867

Note: Accepted model printed in bold. χ^2 values are Satorra–Bentler scaled

Model fit: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

for example by authoritative clinical guidelines (Carr 2014a). This should not only foster the change process in adapting and providing a supportive social context, but also help address the different needs of burdened parents or other family members (Meltzer et al. 2011).

The focus of change in therapy can differ. In family therapy, relational aspects are addressed such as the quality of the relationship, the level of functioning, and the observable interactions between the members of a social system. This might in turn positively influence individual maladaptive behavior, emotion and cognition. Hence, the perception of these affective-cognitive ways of being, interacting, and functioning with each other is of crucial relevance. For instance, youths' perceptions (in this case, of marital relationships) were found to be a predictor for their adjustment succeeding parental ratings (Erel and Kissil 2003). In summary, session monitoring and outcome research should assess all included participants of family therapy.

Research has emphasized the importance of different domains of the relationship between children/adolescents and their parents along development. Fruitful communication and support for each other can evolve when children/adolescents and their parents are affectionately related and when they have feelings of belonging and cohesion embedded in a warm-hearted, caring atmosphere (Paradisopoulos et al. 2015). In addition, an increase of collective functioning is associated with being able to adaptably make decisions about collective aims, find solutions while considering resources – all of these areas are addressed by EVOS (Aguilar-Raab et al. 2018, 2015).

Commonly used self-rating measures of family functioning have rarely been tested for applicability in younger ages. Additionally, it is most often unclear if ratings provided by adolescents are truly comparable to adults' judgments. The present research investigated the applicability of EVOS in a sample of adolescents. Results indicated that EVOS showed comparable psychometric quality in both adolescent and adult samples. Tests for measurement invariance showed that the same two-factor model held across both age groups with only minor deviation. Only a single item ("For me, the way we talk with each other, is …") measuring the communication aspect of a relationship was easier to endorse in younger years. Given that full measurement invariance is a very strict criterion that is often very hard to achieve (Van De Schoot et al. 2015), we still consider this a positive assessment of EVOS' applicability in adolescent samples. The EVOS scale showed equal reliability in both age groups, qualifying EVOS for direct comparisons of observed scores across adult and adolescent samples. Construct validity of EVOS in adolescent samples was established by comparison with SCORE-15, a related measure of family functioning. EVOS and SCORE-15 showed a strong overlap, yet no significant differences

between both measures with regard to criterion validity emerged in our sample. This may be due to the relatively small sample size. One could speculate that the difference between criterion correlations might have turned out to be significant given a larger sample. Such a result would not be surprising, given that SCORE-15 has a problem-focused design, whereas EVOS explicitly does not aim to measure dysfunctional aspects of social relationships. For a closer examination of the relative merits of either scale, much larger samples and ideally multi-trait-multi-method investigations will be necessary (Campbell and Fiske 1959).

The dimensionality of EVOS may still be challenged. EVOS was originally conceived as a measure with two distinct, yet highly related factors. It was consequently constructed using exploratory factor analysis and the two-factor model has been supported by confirmatory factor analysis. Our current data also show that the two-factor model fits the data better than a single-factor model. Nonetheless, both subscales of the EVOS showed similar correlations to other measures. Even though both subscales are conceptually distinct, they are highly correlated and do not show any substantial difference in criterion validity. Going back to the initial scale development (Aguilar-Raab et al. 2015), we have advocated the use of the total score, rather than interpreting the subscales separately. We still argue that EVOS, being already a comprehensive measure based on several theoretical foundations, should be interpreted as a global score, unless a researcher has a specific, theoretically grounded question in mind. Finally, the global score offers the advantage of increased reliability compared to the subscale scores.

Family relationship quality significantly predicted psychological distress among adolescents. Replicating earlier research, our results confirmed the association between family relations and psychological health. The family environment appeared as an important factor for health and well-being (Miller et al. 2009; Rusbult and Van Lange 2003). Prior research has shown that positive social relationships aid the development of self-regulation skills in adolescents (Farley and Kim-Spoon 2014). Contrasting that, negative interactions and low parental involvement have been identified as a risk factor for the development of severe psychopathology (Fruzzetti et al. 2005).

All in all, better family relationship quality was linked to less psychological distress in adolescents. EVOS proved to be an economic, reliable, and valid measure of family relationship quality in adolescents. The present research supports the use of EVOS as an outcome measure in family therapy.

Limitations and Future Research

A potential limitation might be the restricted reliability of the SDQ. The self-report version has been evaluated to some extent and has been successfully used in many languages. Nonetheless the SDQ has shown less than ideal internal consistencies in a large scale evaluation across many countries. Essau et al. (2012) reported internal consistencies for the SDQ total score ranging from Alpha = 0.52 (Italy) to Alpha = 0.74 (Germany). A more recent standardization in Germany reported a slightly higher internal consistency of Alpha = 0.77 for the total score (Lohbeck et al. 2015). In our sample, the SDQ subscale behavioral problems showed only insufficient reliability. This is unfortunate, but still mirrors the known low reliability of the subscale across a wide range of samples (Essau et al. 2012; Lohbeck et al. 2015). We chose this instrument for validation reasons not only because of a lack of available instruments validated in German, but also because the five items of each of the subscales are related but reflect different facets of the characteristic range specified in the scale name, which is particularly desirable for a screening instrument in this context (Lohbeck et al. 2015). Compared to available instruments in the field of research on adolescents, it is a much more economic and well applicable screening instrument for children and youth that can globally assess their behavioral problems and behavioral strengths. In future research, besides examining construct validity by cross-sectional correlational analysis with other more reliable instruments, it is particularly necessary to check the criterion validity in longitudinal studies in order to fully appreciate the advantages of EVOS.

Our analyses of construct validity only included half the sample and only high school students. Unfortunately, this was the result of procedural and ethical concerns on the side of the school administration, yet it does pose a limitation. Future research should investigate EVOS' construct validity in larger probability based samples.

Although we could show measurement invariance of EVOS between different social contexts demonstrating that EVOS assess the same construct of social relationship even in considerably dissimilar situations (Aguilar-Raab et al. 2015), we would like to mention that in our previous studies a majority of adult participants were females. In contrast, in the sub-samples available in the studies presented here, male and female youths were approximately equally matched, with even very slightly more males than females forming the basis of our analysis. Despite the fact that we did not control for or stratified sub-samples by gender, this is a limitation that should be addressed in future research with special focus on gender differences in adolescents in relation to the development of psychopathology and family relationship quality.

Our adolescent sample was from a high socio-economic status associated with a better social background. In a review Conger et al. (2010) pointed to the fact that low socio-economic status has negative (causal) effects on

adults, children and their relationships. Therefore, our findings might not be representative with regard to lower socio-economic status as a result leading to a restricted generalizability. Nonetheless at least with regard to psychological distress our sample was not substantially different from larger, more diverse samples (Lohbeck et al. 2015). Further longitudinal research will be necessary to elucidate causal relations between perceived quality of family relationship as a protective factor and psychological health.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study. The study was approved by the ethics committee of the University Hospital Heidelberg (S-508/2012).

References

Achenbach, T. M. (1991). *Manual for child behavior checklist*. Burlington: University of Vermont.

Achenbach, T. M., & Rescorla, L. A. (2004). The Achenbach System of Empirically Based Assessment (ASEBA) for ages 1.5 to 18 years. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (Vol. 2, 3rd edn.). Mahwah, NJ: Lawrence Erlbaum.

Ackard, D. M., Neumark-Sztainer, D., Story, M., & Perry, C. (2006). Parent-child connectedness and behavioral and emotional health among adolescents. *American Journal of Preventive Medicine*, *30*, 59–66. https://doi.org/10.1016/j.amepre.2005.09.013.

Adelman, C. B., Baucom, D. H., Belus, J. M., Fischer, M. S., & Paprocki, C. (2014). Couple-based interventions for psychopathology: a renewed direction for the field. *Family Process*, *53*, 445–461. https://doi.org/10.1111/famp.12075.

Aguilar-Raab, C., Grevenstein, D., Gotthardt, L., Jarczok, M. N., Hunger, C., Ditzen, B., & Schweitzer, J. (2018). Changing me, changing us: relationship quality and collective efficacy as major outcomes in systemic couple therapy. *Family Process*, *57*, 342–358. https://doi.org/10.1111/famp.12302.

Aguilar-Raab, C., Grevenstein, D., & Schweitzer, J. (2015). Measuring social relationships in different social systems: The construction and validation of the Evaluation of Social Systems (EVOS) Scale. *PLoS ONE*, *10*, e0133442

Allen, J. P., Kilmer, S. L., Kuperminc, G. P., Land, D. J., McElhaney, K. B., Moore, C. W., & O'Beirne–Kelly, H. (2003). A secure base in adolescence: markers of attachment security in the mother–adolescent relationship. *Child Development*, *74*, 292–307. https://doi.org/10.1111/1467-8624.t01-1-00536.

Anderson, N., & West, M. A. (1996). The Team Climate Inventory: Development of the TCI and its applications in teambuilding for innovativeness. *European Journal of Work and Organizational Psychology*, *5*, 53–66. https://doi.org/10.1080/13594329608414840.

Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: an introductory review. *Child and Adolescent Mental Health*, *8*, 136–143. https://doi.org/10.1111/1475-3588.00061.

Atkin, Q. G., Burlingame, G. M., Gawain Wells, M., Lambert, M. J., Latkowski, M., Mosier, J. I., & Whoolery, M. (2001). Tracking the influence of mental health treatment: the development of the youth outcome questionnaire. *Clinical Psychology & Psychotherapy*, *8*, 361–379. https://doi.org/10.1002/cpp.315.

Baldwin, L. M., Bishop, D. S., & Epstein, N. B. (1983). The McMaster family assessment device. *Journal of Marital and Family Therapy*, *9*, 171–180. https://doi.org/10.1111/j.1752-0606.1983.tb01497.x.

Bandura, A., Caprara, G. V., Barbaranelli, C., Regalia, C., & Scabini, E. (2011). Impact of family efficacy beliefs on quality of family functioning and satisfaction with family life. *Applied Psychology*, *60*, 421–448. https://doi.org/10.1111/j.1464-0597.2010.00442.x.

Beavers, W. R. (1981). A systems model of family for family therapists. *Journal of Marital and Family Therapy*, *7*, 299–307. https://doi.org/10.1111/j.1752-0606.1981.tb01382.x.

Beavers, W. R., & Hampson, R. B. (2000). The beavers systems model of family functioning. *Journal of Family Therapy*, *22*, 128–143. https://doi.org/10.1111/1467-6427.00143.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588 https://doi.org/10.1037/0033-2909.88.3.588.

Bishop, D. S., Epstein, N. B., & Levin, S. (1978). The McMaster model of family functioning. *Journal of Marital and Family Therapy*, *4*, 19–31. https://doi.org/10.1111/j.1752-0606.1978.tb00537.x.

Brooks-Gunn, J., Browning, C. R., & Maimon, D. (2010). Collective efficacy, family attachment and urban adolescent suicide attempts. *Journal of Health and Social Behavior*, *51*, 307–324. https://doi.org/10.1177/0022146510377878.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage.

Burlingame, G. M., Cox, J. C., Ferre, R., Lambert, M. J., Latkowski, M., & Wells, M. G. (2005). *Administration and scoring manual for the Y-OQ (Youth Outcome Questionnaire)*. Stevenson, MD: American Professional Credentialing Services.

Burlingame, G. M., Cox, J. C., Lambert, M. J., & Wells, M. G. (2004). *Youth Outcome Questionnaire (Y-OQ)* (Vol. 2, 3rd edn.). Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456 https://doi.org/10.1037/0033-2909.105.3.456.

Calheiros, M. M., & Patrício, J. N. (2014). Assessment of needs in residential care: perspectives of youth and professionals. *Journal of Child and Family Studies*, *23*, 461–474. https://doi.org/10.1007/s10826-012-9702-1.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. https://doi.org/10.1037/h0046016.

Carr, A. (2014a). The evidence base for couple therapy, family therapy and systemic interventions for adult-focused problems. *Journal of Family Therapy*, *36*, 158–194. https://doi.org/10.1111/1467-6427.12033.

Carr, A. (2014b). The evidence base for family therapy and systemic interventions for child-focused problems. *Journal of Family Therapy*, *36*, 107–157. https://doi.org/10.1111/1467-6427.12032.

Carr, A., & Hamilton, E. (2016). Systematic review of self-report family assessment measures. *Family Process*, *55*, 16–30. https://doi.org/10.1111/famp.12200.

Carr, A., & Stratton, P. (2017). The score family assessment questionnaire: a decade of progress. *Family Process*, *56*, 285–301. https://doi.org/10.1111/famp.12280.

Celinska, K., Cheng, C.-C., & Virgil, N. J. (2015). Youth and parental perspectives on the functional family therapy programme. *Journal of Family Therapy*, *37*, 450–470. https://doi.org/10.1111/1467-6427.12051.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. https://doi.org/10.1080/10705510701301834.

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005–1018. https://doi.org/10.1037/a0013193.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233–255. https://doi.org/10.1207/S15328007SEM0902_5.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*, 347–357. https://doi.org/10.1111/j.2044-8317.1991.tb00966.x.

Cleveland, K., Garrett, S. B., Hair, E. C., Ling, T., & Moore, K. A. (2008). The continued importance of quality parent–adolescent relationships during late adolescence. *Journal of Research on Adolescence*, *18*, 187–200. https://doi.org/10.1111/j.1532-7795.2008.00556.x.

Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes, and individual development. *Journal of Marriage and the Family*, *72*, 685–704. https://doi.org/10.1111/j.1741-3737.2010.00725.x.

Cook, T. D., Herman, M. R., Phillips, M., & Settersten, J. R. A. (2002). Some ways in which neighborhoods, nuclear families, friendship groups, and schools jointly affect changes in early adolescent development. *Child Development*, *73*, 1283–1309. https://doi.org/10.1111/1467-8624.00472.

Demaray, M. K., & Malecki, C. K. (2002). The relationship between perceived social support and maladjustment for students at risk. *Psychology in the Schools*, *39*, 305–316. https://doi.org/10.1002/pits.10018.

Erel, O., & Kissil, K. (2003). The linkage between multiple perspectives of the marital relationship and preschoolers' adjustment. *Journal of Child and Family Studies*, *12*, 411–423. https://doi.org/10.1023/a:1026012023182.

Essau, C. A., Olaya, B., Anastassiou-Hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D., & Ollendick, T. H. (2012). Psychometric properties of the strength and difficulties questionnaire from five European countries. *International Journal of Methods in Psychiatric Research*, *21*, 232–245. https://doi.org/10.1002/mpr.1364.

Farley, J. P., & Kim-Spoon, J. (2014). The development of adolescent self-regulation: reviewing the role of parent, peer, friend, and romantic relationships. *Journal of Adolescence*, *37*, 433–440. https://doi.org/10.1016/j.adolescence.2014.03.009.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146.

Faulstich, P. (1998). *Strategien betrieblicher Weiterbildung. Kompetenz und Organisation*. München, Germany: Vahlen.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*, 39–50. https://doi.org/10.2307/3151312.

Fruzzetti, A. E., Shenk, C., & Hoffman, P. D. (2005). Family interaction and the development of borderline personality disorder: a transactional model. *Development and Psychopathology*, *17*, 1007–1030. https://doi.org/10.1017/S0954579405050479.

Goodman, R. (1997). The Strengths and difficulties questionnaire: a research note. *Child Psychology & Psychiatry & Allied Disciplines*, *38*, 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

Grevenstein, D., Bluemke, M., Schweitzer, J., & Aguilar-Raab, C. (2018). *Better family relationships - higher well-being: The connection between relationship quality and health related ressources*. Manuscript submitted for publication.

Hamilton, E., Carr, A., Cahill, P., Cassells, C., & Hartnett, D. (2015). Psychometric properties and responsiveness to change of 15-and 28-Item versions of the SCORE: A family assessment questionnaire. *Family Process*, *54*, 454–463. https://doi.org/10.1111/famp.12117.

Heatherington, L., Diamond, G. M., Escudero, V., Friedlander, M. L., & Pinsof, W. M. (2015). 25 Years of systemic therapies research: Progress and promise. *Psychotherapy Research*, *25*, 348–364. https://doi.org/10.1080/10503307.2014.983208.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/10705519909540118.

Johnson, H. D., Lavoie, J. C., & Mahoney, M. (2001). Interparental conflict and family cohesion. *Journal of Adolescent Research*, *16*, 304–318. https://doi.org/10.1177/0743558401163004.

Jose, P. E., Ryan, N., & Pryor, J. (2012). Does social connectedness promote a greater sense of well-being in adolescence over time? *Journal of Research on Adolescence*, *22*, 235–251. https://doi.org/10.1111/j.1532-7795.2012.00783.x.

Kauffeld, S. (2001). *Teamdiagnose*. Göttingen, Germany: Hogrefe.

Kenny, D. A. (2015). Measuring model fit. Retrieved from http://davidakenny.net/cm/fit.htm

Laursen, B., & Colling, W. A. (2004). Chapter 16: Parent-child communication during adolescence. In A. L. Vangelisti (Ed.), *Handbook of family communication*. Mahwah, NJ: Lawrence Erlbaum.

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common. Retrieved from http://www.quantpsy.org

Lohbeck, A., Schultheiß, J., Petermann, F., & Petermann, U. (2015). Die deutsche selbstbeurteilungsversion des strengths and difficulties questionnaire (SDQ-Deu-S). *Diagnostica*, *61*, 222–235. https://doi.org/10.1026/0012-1924/a000153.

Lubke, G. H., & Dolan, C. V. (2003). Can unequal residual variances across groups mask differences in residual means in the common factor model? *Structural Equation Modeling*, *10*, 175–192. https://doi.org/10.1207/S15328007SEM1002_1.

Lucia, V. C., & Breslau, N. (2006). Family cohesion and children's behavior problems: a longitudinal investigation. *Psychiatry Research*, *141*, 141–149. https://doi.org/10.1016/j.psychres.2005.06.009.

Mansfield, A. K., Keitner, G. I., & Dealy, J. (2015). The family assessment device: an update. *Family Process*, 54, 82–93. https://doi.org/10.1111/famp.12080.

McClendon, D. T., Warren, J. S., M. Green, K., Burlingame, G. M., Eggett, D. L., & McClendon, R. J. (2011). Sensitivity to change of youth treatment outcome measures: a comparison of the CBCL, BASC-2, and Y-OQ. *Journal of Clinical Psychology*, 67, 111–125. https://doi.org/10.1002/jclp.20746.

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: past, present, and future. *Personality and Social Psychology Review*, 4, 95–105. https://doi.org/10.1207/S15327957PSPR0401_8.

Meltzer, H., Ford, T., Goodman, R., & Vostanis, P. (2011). The burden of caring for children with emotional or conduct disorders. *International Journal of Family Medicine*, 2011. https://doi.org/10.1155/2011/801203

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. https://doi.org/10.1007/BF02294825.

Merikangas, K. R., He, J.-P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication - Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49, 980–989. https://doi.org/10.1016/j.jaac.2010.05.017.

Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*, 11, 7–20. https://doi.org/10.1007/978-3-319-25501-9_1.

Miller, G., Chen, E., & Cole, S. W. (2009). Health psychology: developing biologically plausible models linking the social world and physical health. *Annual Review of Psychology*, 60, 501–524. https://doi.org/10.1146/annurev.psych.60.110707.163551.

Moore, K. A., Guzman, L., Hair, E. C., Lippman, L., & Garrett, S. B. (2004). *Parent–teen relationship and interactions: Far more positive than not*. Washington, DC: Child Trends.

Moore, L., & Seu, I. B. (2011). Giving children a voice: children's positioning in family therapy. *Journal of Family Therapy*, 33, 279–301. https://doi.org/10.1111/j.1467-6427.2011.00556.x.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus*. Los Angeles, CA: Muthén & Muthén. 2012.

Noller, P., Feeney, J., & Petersen, C. (2001). *Personal relationships across the lifespan*. London, UK: Psychology Press.

O'Brien, M. P., Gordon, J. L., Bearden, C. E., Lopez, S. R., Kopelowicz, A., & Cannon, T. D. (2006). Positive family environment predicts improvement in symptoms and social functioning among adolescents at imminent risk for onset of psychosis. *Schizophrenia Research*, 81, 269–275. https://doi.org/10.1016/j.schres.2005.10.005.

Olson, D. H. (1985). *FACES III - Family Adaptability and Cohesion Evaluation Scales*. St. Paul, MN: University of Minnesota.

Olson, D. H. (2011). FACES IV and the circumplex model: validation study. *Journal of Marital and Family Therapy*, 37, 64–80. https://doi.org/10.1111/j.1752-0606.2009.00175.x.

Olson, D. H., Sprenkle, D. H., & Russell, C. S. (1979). Circumplex model of marital and family systems: I. cohesion and adaptability dimensions, family types, and clinical applications. *Family Process*, 18, 3–28. https://doi.org/10.1111/j.1545-5300.1979.00003.x.

Paradisopoulos, D., Pote, H., Fox, S., & Kaur, P. (2015). Developing a model of sustained change following multisystemic therapy: young people's perspectives. *Journal of Family Therapy*, 37, 471–491. https://doi.org/10.1111/1467-6427.12070.

Patterson, J. M. (2002). Understanding family resilience. *Journal of Clinical Psychology*, 58, 233–246. https://doi.org/10.1002/jclp.10019.

Phillips-Salimi, C. R., Robb, S. L., Monahan, P. O., Dossey, A., & Haase, J. E. (2014). Perceptions of communication, family adaptability and cohesion: a comparison of adolescents newly diagnosed with cancer and their parents. *International Journal of Adolescent Medicine & Health*, 26, 19–26. https://doi.org/10.1515/ijamh-2012-0105.

Pinquart, M., Oslejsek, B., & Teubert, D. (2016). Efficacy of systemic therapy on adults with mental disorders: a meta-analysis. *Psychotherapy Research*, 26, 241–257. https://doi.org/10.1080/10503307.2014.935830.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. https://doi.org/10.2307/271063.

Rayner, M., & Montague, M. (2000). *Resilient children and young people*. Melbourne: Deakin University.

Resnick, M. D., Harris, L. J., & Blum, R. W. (1993). The impact of caring and connectedness on adolescent health and well-being. *Journal of Paediatrics and Child Health*, 29, S3–S9. https://doi.org/10.1111/j.1440-1754.1993.tb02257.x.

Retzlaff, R., von Sydow, K., Beher, S., Haun, M. W., & Schweitzer, J. (2013). The efficacy of systemic therapy for internalizing and other disorders of childhood and adolescence: a systematic review of 38 randomized trials. *Family Process*, 52, 619–652. https://doi.org/10.1111/famp.12041.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavioral assessment system for children: Second edition (BASC-2)*. Circle Pines, MN: AGS Publishing.

Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology*, 54, 351–375. https://doi.org/10.1146/annurev.psych.54.101601.145059.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in multivariate statistical analysis. A Festschrift for Heinz Neudecker* (pp. 233–247). London, UK: Kluwer Academic Publishers.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. https://doi.org/10.1007/BF02296192.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: review of practice and implications. *Human Resource Management Review*, 18, 210–222. https://doi.org/10.1016/j.hrmr.2008.03.003.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464. https://doi.org/10.1214/aos/1176344136.

Sexton, T. L., Datchi, C., Evans, L., LaFollette, J., & Wright, L. (2013). The effectiveness of couple and family-based clinical interventions. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*. (6th edn., pp. 587–639). Hoboken, NJ: John Wiley & Sons.

Skinner, H. A., Steinhauer, P. D., & Santa-Barbara, J. (1983). The family assessment measure. *Canadian Journal of Community Mental Health*, 2, 91–103. https://doi.org/10.1111/1467-6427.00146.

Skinner, H. A., Steinhauer, P. D., & Sitarenios, G. (2000). Family assessment measure (FAM) and process model of family functioning. *Journal of Family Therapy*, 22, 190–210. https://doi.org/10.1111/1467-6427.00146.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245–251. https://doi.org/10.1037/0033-2909.87.2.245.

Stratton, P., Bland, J., Janes, E., & Lask, J. (2010). Developing an indicator of family function and a practicable outcome measure for systemic family and couple therapy: the SCORE. *Journal of Family Therapy*, *32*, 232–258. https://doi.org/10.1111/j.1467-6427.2010.00507.x.

Stratton, P., Lask, J., Bland, J., Nowotny, E., Evans, C., Singh, R., & Peppiatt, A. (2014). Detecting therapeutic improvement early in therapy: validation of the SCORE-15 index of family functioning and change. *Journal of Family Therapy*, *36*, 3–19. https://doi.org/10.1111/1467-6427.12022.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement Invariance. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2015.01064

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. https://doi.org/10.1177/109442810031002.

von Sydow, K., Retzlaff, R., Beher, S., Haun, M. W., & Schweitzer, J. (2013). The efficacy of systemic therapy for childhood and adolescent externalizing disorders: a systematic review of 47 RCT. *Family Process*, *52*, 576–618. https://doi.org/10.1111/famp.12047.

Walsh, F. (2003). Family resilience: a framework for clinical practice. *Family Process*, *42*, 1–18. https://doi.org/10.1111/j.1545-5300.2003.00001.x.

Wells, M. G., Burlingame, G. M., & Rose, P. M. (2003). *Administration and scoring manual for the Y-OQ SR-2.0 (Youth Outcome Questionnaire-Self Report)*. Wilmington, DE: American Professional Credentialing Services.

Wells, M. S., Widmer, M. A., & McCoy, J. K. (2004). Grubs and grasshoppers: challenge-based recreation and the collective efficacy of families with at-risk youth. *Family Relations*, *53*, 326–333. https://doi.org/10.1111/j.0197-6664.2003.0009.x.

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200. https://doi.org/10.1111/0081-1750.00078.