



OPEN ACCESS

EDITED BY

Cesar Merino-Soto,
Monterrey Institute of Technology and
Higher Education (ITESM), Mexico

REVIEWED BY

Nicola Lovecchio,
University of Bergamo, Italy
Aikaterini Vasiou,
University of Crete, Greece

*CORRESPONDENCE

Dirk Baier
✉ dirk.baier@zhaw.ch

RECEIVED 17 March 2025
REVISED 03 February 2026
ACCEPTED 04 February 2026
PUBLISHED 27 February 2026

CITATION

Schmalbach B, Baier D, Krieg Y,
Brähler E and Kliem S (2026)
Construction and validation of the
ultra-short version of the Parenting
Scale (PS-4).
Front. Psychol. 17:1595258.
doi: 10.3389/fpsyg.2026.1595258

COPYRIGHT

© 2026 Schmalbach, Baier, Krieg,
Brähler and Kliem. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Construction and validation of the ultra-short version of the Parenting Scale (PS-4)

Bjarne Schmalbach¹, Dirk Baier^{2,3*}, Yvonne Krieg⁴,
Elmar Brähler^{5,6} and Sören Kliem⁷

¹Medical Psychology and Medical Sociology, University Medical Center of the Johannes Gutenberg University, Mainz, Germany, ²Institute of Delinquency and Crime Prevention, Zurich University of Applied Sciences, Zurich, Switzerland, ³Faculty of Law, University of Zurich, Zürich, Switzerland, ⁴Criminological Research Institute of Lower Saxony, Hannover, Germany, ⁵Clinic and Polyclinic for Psychosomatic Medicine and Psychotherapy, Rostock University Medical Center, Rostock, Germany, ⁶Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, ⁷Ernst-Abbe-Hochschule Jena – University of Applied Sciences, Jena, Germany

Parenting behavior is a central determinant of childhood development and is thus deserving of more scientific attention. In the present article, we constructed an ultra-short scale for the assessment of parenting styles, the Parenting Scale 4 (PS-4). To this end, we analyzed large samples of parent–child dyads—one representative of the German general population (Sample 1), the other representative for the German federal state Lower Saxony (Sample 2). We applied an algorithm-based scale-shortening technique in Sample 1 and confirmed the resulting model in Sample 2, finding excellent model fit and—given the extreme brevity—acceptable reliability. Furthermore, we show that the model is invariant across parent and child genders. Correlations with the Strengths and Difficulties Questionnaire remain virtually unchanged compared to a longer version of the Parenting Scale, which is evidence for the PS-4's validity. Overall, the PS-4 can be recommended for the assessment of parenting behavior, particularly in large-scale surveys with time constraints.

KEYWORDS

assessment, child development, mental health, parenting behavior, scale development

Introduction

Parenting is one of the most important determinants of childhood development (Belsky, 1984; Sulik et al., 2015; Tran et al., 2017). A myriad of crucial outcomes is shaped by the way parents treat and teach their children: health, life satisfaction, and success later in life among many others (Abubakar et al., 2015; Hamilton, 2016; McLeod and Shanahan, 1993; Milevsky et al., 2007; Raboteg-Saric and Sakic, 2014; Steinberg et al., 1989; Waters, 2015).

The Parenting Scale (PS; Arnold et al., 1993) is one of the most widely applied parenting behavior assessment tools - also in the evaluation of trainings and prevention programs (de Graaf et al., 2008; Salari et al., 2012). The original PS consists of 30 items and can be divided into two factors: Overreactivity and Laxness. A third factor (Verbosity) was discussed in the beginning but has been refuted by subsequent research (for a review see Salari et al., 2012). Various shortforms of the PS have been developed over the years (Lorber et al., 2014; Reitman et al., 2001; Rhoades and O'Leary, 2007). The shortest that we are aware of is the 8-item version by Kliem et al. (2019). While eight items is very brief, even shorter scales—such as the Patient Health Questionnaire 4 (Kroenke et al., 2009; Löwe et al., 2010)—have been developed and

enjoy great popularity. They are particularly useful in survey research, where every item counts and minimizing time investment is of the utmost importance.

The present study thus served the purpose of shortening the PS-8 to the ultra-short screening version PS-4. To this end, we first examined the psychometric properties of the PS-8 items and constructed a subset of 2 items per scale. We then tested this subset in a second, independent sample using confirmatory factor analysis (CFA). In addition, we tested for measurement invariance between genders and tested correlations with the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). Since the PS is supposed to capture problematic parenting tendencies, we expect positive correlations between the PS-4 subscales and the difficulties subscales of the SDQ and positive correlations with the strengths subscale.

Method

Procedure and participants

Sample 1 was collected as part of a representative survey of the German general population in 2013. The demographic consulting company (USUMA, Berlin, Germany) which was responsible for data collection, initially contacted a total of 4,360 individuals based on random selection. The selection procedure uses three steps: 258 geographic regions in Germany, households within a specific region, and individuals in a specific household. Of those, many refused to provide any information ($n = 591$), were not present upon visitation ($n = 563$), refused to be interviewed ($n = 540$), or were not included for other pre-defined reasons ($n = 158$). A trained interviewer visited all participants. After making sure that they possessed sufficient command of the German language, the interviewer informed them of the general purpose of the study and obtained informed consent. In the case of minors, consent of the parents was also obtained. Following a sociodemographic interview, the participants completed self-report questionnaires in the presence of (but without any interference from) the interviewer. The final study sample consisted of $n = 2,508$ individuals. Out of those, 539 reported currently raising a child. Eight of these observations were excluded because of missing values. Thus, our exploratory analysis sample consisted of 531 observations. No systematic differences between participants with complete and incomplete data were observed. Given the low and largely unsystematic proportion of missingness, no imputation procedures were applied. The study was authorized by the Ethics Committee of the Medical Faculty of the University of Leipzig (Az.: 050/13-03.05.2013).

Sample 2, was collected as part of a representative survey of fourth graders in the German federal state of Lower Saxony, carried out by the Criminological Research Institute of Lower Saxony (Kriminologisches Forschungsinstitut Niedersachsen – KFN) as part of an evaluation study of the “Klasse2000” universal prevention program (Kliem et al., 2020a,b) and was funded by the Federal Centre for Health Education (Bundeszentrale für gesundheitliche Aufklärung – BZgA). The survey involved a random sample of year four classrooms in Lower Saxony from all classrooms taught in the 2016/2017 school year. In total, $K = 740$ classes were selected for the survey. The final sample was reduced to $k = 528$ classrooms based on refusals to participate from school principals or teachers; this corresponds to a classroom participation rate of 71.4%. Of the $N = 9,880$ students attending

the participating classes, $n = 6,376$ students participated in the survey, which corresponds to a student participation rate of 64.5%. All 6,376 participating children received a questionnaire for their parents by “satchel post.” From 3,324 of these parents a completed questionnaire was sent back to the KFN. This corresponds to a response rate of 52.1%. In addition, we removed observations with missing values. Thus, our final sample consists of $n = 3,133$ observations. Only children whose parents returned a signed consent form to the teacher were allowed to take part in the interview. The ethics committee of the Georg Elias Müller Institute for Psychology at the University of Göttingen (NO. 184/2018) has given a positive ethics statement for this procedure. Furthermore, the survey was approved by the regional school authorities (Landesschulbehörde Niedersachsen H1Rb-81402-09-2017). During the entire survey, the children were guided through the eleven-page questionnaire by the test leaders using a manual. Participation is entirely voluntary among principals, teachers, parents, and children, and nonparticipation does not lead to any negative consequences. Children whose parents have not provided written informed consent are not surveyed. Teachers administered the written informed consent forms, did not forward them to the researchers, and destroyed them 2 weeks after the survey was conducted. Since the study design provides for complete anonymization, no signature of the participating adults could be obtained for their own participation. The sociodemographic variables of both samples are displayed in Table 1.

Instruments

The Parenting Scale 8 (PS-8, Kliem et al., 2019) consists of two subscales - overreactivity and laxness - measured with four items each. Respondents describe their typical reactions to certain behaviors a child can exhibit. All items are bi-polar, giving two potential extremes of the reaction spectrum. Respondents rate on a scale of 1 to 7 to what extent these responses are typical of them. Reliability coefficients for the present study are reported in Table 2.

Sample 2 also included the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001; Klasen et al., 2003). This questionnaire consists of 25 items that assess a parent’s view of their child’s particular strengths and weaknesses in terms of behavior, self-regulation, and social behavior. There are five subscales: Conduct Problems,

TABLE 1 Sociodemographic variables of both samples.

	Sample 1 ($n = 531$)		Sample 2 ($n = 3,133$)	
	n	%	n	%
Parent gender				
Male	220	41.4	345	11.1
Female	311	58.6	2,776	88.9
Parent age				
M, SD	39.56	8.30	42.08	5.54
Child gender				
Male	278	52.7	1,531	49.0
Female	250	47.3	1,592	51.0
Child age				
M, SD	8.92	5.23	2.69	0.58

TABLE 2 Correlation matrix of the Parenting Scales 4 and 8 as well as the strengths and difficulties questionnaire (Sample 2).

	1	2	3	4	5	6	7	8	9
1. Overreactivity 4	0.622								
2. Laxness 4	0.329	0.546							
3. Overreactivity 8	0.847/0.790*	0.271	0.645						
4. Laxness 8	0.387	0.895/0.854*	0.326	0.714					
5. Emotional Symptoms (SDQ)	0.165	0.093	0.164	0.109	0.696				
6. Conduct Problems (SDQ)	0.283	0.155	0.312	0.176	0.295	0.554			
7. Hyperactivity (SDQ)	0.209	0.112	0.244	0.137	0.329	0.498	0.791		
8. Peer Problems (SDQ)	0.147	0.107	0.114	0.138	0.383	0.307	0.291	0.676	
9. Prosocial Behavior (SDQ)	-0.203	-0.104	-0.233	-0.138	-0.098	-0.400	-0.277	-0.176	0.660

Internal consistency is displayed on the diagonal, *adjusted for overlap in accordance with Levy (1967).

Emotional Symptoms, Hyperactivity, Peer Problems, and Prosocial Behavior. Response options range from 1 (*Not true*) to 3 (*Certainly true*). Reliability coefficients are reported in Table 2.

Statistical analyses

We conducted all analyses in R, using the *EFAutilities*, *ezCutoffs*, *lavaan*, *matrixStats*, *moments*, *multilevel*, *paran*, *semTools*, and *stuart* packages (Bengtsson, 2020; Bliese, 2016; Dinno, 2018; Jorgensen et al., 2020; Komsta and Novomestky, 2015; Rosseel, 2012; Schmalbach et al., 2019; Schultze, 2020; Zhang et al., 2020). We excluded those respondents from the analyses who had not answered the PS-8 scale or had missing data on one or more of the items. We applied exploratory factor analysis and parallel analysis (PA; Horn, 1965) to Sample 1. The purpose of these procedures was to check how many components should reasonably be extracted from the eight indicators, and how the items load on them. We consider loadings of 0.50 and greater as sufficient. In addition, there should be no substantial cross-loadings (0.20). After this initial check we used *stuart's* *bruteforce* algorithm to find the best 2 × 2 solution among all possible item subsets. “Best solution” here means best model fit and best composite reliability.

We then fitted the resulting model in Sample 2 using robust diagonally weighted squares estimation Li (2016) and delta parameterization. In addition to the χ^2 -test, we utilized the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Squared Error of Approximation (RMSEA), and the Standardized Root Mean Squared Residual (SRMR) to evaluate model fit. We judged these indices based on the traditionally recommended cutoff values: the p -value of the χ^2 -test should be greater than 0.05 (0.01), CFI/TLI greater than 0.97 (0.95), RMSEA smaller than 0.05 (0.08), and SRMR smaller than 0.05 (0.10), respectively, to indicate good (acceptable) fit between theoretical model and empirical data (Schermelleh-Engel et al., 2003). To supplement these traditional fixed cutoff values, we employed the simulation-based approach introduced by Schmalbach et al. (2019). Here, the empirically derived fit values are compared to the fit index distributions obtained by simulating a large number of data sets with the same underlying parameter table and overall same properties as the empirical one. As generally recommended, we estimate reliability using McDonald's ω (Dunn et al., 2014; Green and Yang, 2009, Formula 21). To test for measurement invariance between genders, we followed the guidelines provided by Wu and Estabrook (2016) and evaluated the comparisons using the recommended procedure (Chen, 2007). That is, we successively constrained item thresholds, factor loadings, item intercepts, and item residual variances to be equal

between groups. For each of these steps $\Delta\chi^2$ should be non-significant, ΔCFI should be smaller than 0.01 and $\Delta RMSEA$ should be smaller than 0.015.

Results

Item descriptive statistics and exploratory analyses (Sample 1)

We report item descriptive statistics in Table 3. All items were approximately normally distributed upon inspection of skewness and kurtosis (Kim, 2013). Parallel analysis showed clear evidence for two substantial components in the PS-8. The first three empirical eigenvalues were 2.62, 1.57, and 0.72 - compared to 1.24, 1.16, and 1.10 in the randomly generated data sets. Loadings from the subsequent exploratory factor analysis are also reported in Table 3. It should be noted that Items 5 and 8 had comparatively low loadings. In addition, Item 8 evinced a cross-loading. Next, we entered all items into the *stuart* to find the optimal two-factor solution among the possible 36 subsets. The best solution, which had excellent model fit, $\chi^2(1) = 0.667$, $p = 0.414$, CFI = 1, TLI = 1.006, RMSEA = 0, SRMR = 0.006, is marked in Table 3.

Confirmatory analyses (Sample 2)

We then tested the model generated in the EFA sample in the independent CFA sample. As can be seen in Table 4, fit measures indicated very good fit for the two-factor model and the χ^2 -test was not significant at $p = 0.134$. Even the simulated cutoffs, which are in this instance much stricter than the traditional cutoffs, were not violated. The two latent variables correlated strongly, $r = 0.581$, $p < 0.001$. Reliability was $\omega_{\text{Overreactivity}} = 0.622$ [0.601, 0.643] and $\omega_{\text{Laxness}} = 0.546$ [0.522, 0.568], respectively. In general, these coefficients would be considered small, even unacceptable. However, it should be noted that ultra-short scales tend to yield smaller reliability coefficients. Based on the Spearman-Brown prophecy formula, we can estimate that two four-item scales of the same general characteristics would evince ω s of 0.767 and 0.706, respectively. In order to confirm the two-dimensionality of the PS-4 scale, we then also tested the uni-dimensional version of the model. However, this model evinced unacceptable fit by all accounts.

TABLE 3 Item descriptive statistics of the Parenting Scale 8 (Sample 1).

	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	λ F1	λ F2
Item PS1 ^O	5.30	1.48	-0.66	2.58	-0.036	0.688
Item PS2 ^{O,4}	5.79	1.18	-1.05	3.90	0.092	0.737
Item PS3 ^{L,4}	5.19	1.45	-0.67	2.98	0.763	0.055
Item PS4 ^L	5.13	1.51	-0.78	3.06	0.653	-0.043
Item PS5 ^{R,O}	4.33	1.75	-0.16	2.00	-0.107	0.484
Item PS6 ^{L,4}	5.66	1.24	-1.08	4.18	0.769	0.030
Item PS7 ^{R,O,4}	5.82	1.43	-1.34	4.31	0.064	0.536
Item PS8 ^{R,L}	5.57	1.35	-0.94	3.58	0.466	0.240

M, Mean; *SD*, Standard Deviation; λ , Standardized factor loadings on factors 1 and 2 in the exploratory factor analysis; R, Reverse-coded item; O, Overreactivity scale; L, Laxness scale; 4, Part of the PS-4.

TABLE 4 Model fit statistics of the Parenting Scale 4 (Sample 2).

	χ^2	CFI	TLI	RMSEA	SRMR
2FM empirical fit	2.247	1.000	0.997	0.020	0.006
2FM simulated cutoffs	3.984	0.998	0.990	0.031	0.009
1FM empirical fit	242.516	0.914	0.741	0.196	0.060
1FM simulated cutoffs	5.830	0.998	0.994	0.025	0.011

Instances in which the empirical fit measure is worse than the simulated cutoff should be considered evidence of model misfit.

Next, we tested for measurement invariance between genders. This means that we considered both the child's and the parent's gender as well as the match/mismatch between the two. The results for these analyses are summarized in Table 1. As is evident from the non-significant increases in χ^2 - in addition to the virtually unchanged values for *CFI* and *RMSEA* - from step to step, the model can be considered to be invariant across groups of child gender, parent gender, and child-parent gender match/mismatch. Merely the *RMSEA* for the configural model of parent gender is substantially higher than in the more constrained models. This can probably be explained by the low number of free parameters, which is known to inflate *RMSEA* (Kenny et al., 2015). Overall, however, we find strong evidence for gender invariance. This is an important requisite for comparisons of both latent and observed means and variances between these populations (Gregorich, 2006).

Correlations with the SDQ

Overall, we observed the expected pattern of correlations between the PS-4/-8 and the SDQ scales (see Table 2). Specifically, there were small to moderate positive correlations between the two subscales of the PS-4/-8 and the four problem-oriented scales of the SDQ. In contrast, the SDQ prosocial behavior scale correlated negatively with both PS subscales. In order to gauge the difference between the 8-item PS and the ultra-short PS-4, we first examined the correlations between the two versions of the scales. As is evident with the very high correlations, the majority of the PS-8's variance is still explained by the PS-4. In investigating the differences between correlations of the long and short versions of the PS and the SDQ subscales this similarity becomes even clearer: For Overreactivity, the average difference was $\Delta z > -0.001$, $p = 0.999$, while for Laxness the average difference was $\Delta z = -0.012$, $p = 0.992$. This means, one can use half the items and apply the PS-4 instead of the PS-8 with a negligible loss of information and validity.

Discussion

The aim of the research at hand was the construction of an ultra-short version of the PS. This was achieved by drawing on two large samples - one representative of the German general population, the other representative for the German federal state of Lower Saxony.

While a loss of reliability is close to inevitable while eliminating items, we observed exactly the decrease that was to be expected for a reduction from four to two items per scale. Model fit, on the other hand, improved greatly. All criteria evinced excellent goodness-of-fit. Furthermore, we found evidence for measurement invariance between genders: This means that the observed means can be compared both between different gendered children and parents. In addition, we examined whether the matching of same /different gendered children and parents could result in a lack of invariance, which was not the case. As is evident with the correlation matrix of the PS-4, PS-8, and SDQ, information loss was minimal in eliminating four items of the PS-8. Specifically, we observe the same pattern of correlations and correlational magnitudes for both the 4- and 8-item version: Positive correlations for the Prosocial Behavior scale and negative correlations for the other four scales, which deal with problematic tendencies.

Limitations

The present study is limited in that we have not applied the final four PS-4 items in isolation. It appears unlikely that the outcomes of the psychometric investigation should change, but future studies should strive to replicate our findings to be sure.

As in all survey studies, the data may have been falsified by the respondents due to processes of social desirability or deliberate deception, whereby such processes can be avoided by informing the participants in detail about the underlying data protection mechanisms (e.g., anonymous survey as well as in sample 2 no feedback to

parents, school or teachers) should be kept to a minimum. Response rates were = 57.5%/57.6%, which is quite common in general population studies (e.g., Kliem et al., 2014, 2015, 2018, 2021, 2024, 2025; Schmalbach et al., 2019) as well as in school surveys (Beckmann et al., 2021; Kliem et al., 2020a,b; Rehbein et al., 2015). However, it cannot be excluded that participants/students with a high level of stress often did not participate in the surveys. A further limitation concerns potential partner and co-parent influences in families with separated parents. Although the study focuses on the primary caregiver, child outcomes are likely shaped by complex interactions with non-resident parents and new partners (e.g., co-parenting quality, conflict, parenting practices in the other household). These influences could not be reliably measured or disentangled in the present design and may have introduced unobserved heterogeneity. Future studies should incorporate multi-informant and multi-household assessments to better capture interparental dynamics and their impact on child outcomes. Furthermore, Germany, as well as most other countries in northern Europe, has abolished any form of parental violence (physical as well as emotional, i.e., corporal punishment) as an accepted form of parental educational method. As of the year 2000 corporal punishment is considered abusive and hence a criminal offence under German law (§ 1,631 Abs. 2 *Bürgerliches Gesetzbuch* [German Civil Code]) which is not the case in several US states. It can therefore be assumed that German parents are more prone to giving socially desirable answers than their US counterparts. Parenting behaviors and disciplinary practices are embedded within cultural, legal, and normative contexts. The PS-4 was developed and validated using samples from Western, high-income countries, where parenting norms and legal regulations surrounding child-rearing are relatively similar. Its generalizability to cultural contexts with different normative parenting practices, legal frameworks, or social expectations cannot therefore be assumed. Future research should explicitly test cross-cultural measurement invariance and cultural validity before applying the PS-4 in non-Western or legally divergent contexts.

Lastly, the ultra-short format of the PS-4 entails a trade-off between measurement efficiency and content coverage. While the scale captures core dimensions of dysfunctional parenting, it cannot fully reflect the behavioral breadth and nuance assessed by longer versions of the Parenting Scale. Accordingly, the PS-4 is not intended to replace longer instruments such as the PS-8, nor is it suitable for individual-level assessment or clinical decision-making. Rather, it should be understood as an ultra-short screening instrument designed for contexts in which assessment time and respondent burden must be minimized, such as large-scale surveys and population-based research. Although the PS-4 demonstrates good model fit and evidence of construct validity given its minimal item count, longer versions remain preferable when more detailed assessment is required. Due to its limited reliability, associations involving the PS-4 may be attenuated and should be interpreted with appropriate caution.

Conclusion

In sum, the PS-4 can be recommended as an ultra-short, yet valid and (given its extreme brevity) reliable assessment tool of parenting behavior. It is particularly useful in large-scale surveys where time is of the essence.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Medical Faculty of the University of Leipzig Ethics committee of the Georg Elias Müller Institute for Psychology at the University of Göttingen. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

BS: Writing – original draft, Formal analysis, Methodology, Data curation. DB: Writing – review & editing. YK: Writing – review & editing. EB: Writing – review & editing, Project administration. SK: Writing – review & editing, Conceptualization, Project administration.

Funding

The author(s) declared that financial support was received for this work and/or its publication. Open access funding by Zurich University of Applied Sciences (ZHAW).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author SK declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Abubakar, A., Van de Vijver, F. J. R., Suryani, A. O., Handayani, P., and Pandia, W. S. (2015). Perceptions of parenting styles and their associations with mental health and life satisfaction among urban Indonesian adolescents. *J. Child Fam. Stud.* 24, 2680–2692. doi: 10.1007/s10826-014-0070-x
- Arnold, D. S., O'Leary, S. G., Wolff, L. S., and Acker, M. M. (1993). The parenting scale: a measure of dysfunctional parenting in discipline situations. *Psychol. Assess.* 5, 137–144. doi: 10.1037/1040-3590.5.2.137
- Beckmann, L., Bergmann, M. C., Krieg, Y., and Kliem, S. (2021). Associations between classroom normative climate and the perpetration of teen dating violence among secondary school students. *J. Interpers. Violence* 36:NP11291–NP11321. doi: 10.1177/0886260519888207
- Belsky, J. (1984). The determinants of parenting: a process model. *Child Dev.* 55, 83–96. doi: 10.1111/j.1467-8624.1984.tb00275.x
- Bengtsson, H. (2020). matrixStats: functions that apply to rows and columns of matrices (and to vectors). R package version 0.57.0. Available online at: <https://CRAN.R-project.org/package=matrixStats> (Accessed February 16, 2026).
- Bliese, P. (2016). multilevel: multilevel functions. R package version 2.6. Available online at: <https://CRAN.R-project.org/package=multilevel> (Accessed February 16, 2026).
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equ. Model. Multidiscip. J.* 14, 464–504. doi: 10.1080/10705510701301834
- de Graaf, I., Speetjens, P., Smit, F., de Wolff, M., and Tavecchio, L. (2008). Effectiveness of the triple p positive parenting program on behavioral problems in children: a meta-analysis. *Behav. Modif.* 32, 714–735. doi: 10.1177/0145445508317134
- Dinno, A. (2018). Paran: horn's test of principal components/factors (R package version 1.4.2). Available online at: <https://CRAN.R-project.org/package=paran> (Accessed February 16, 2026).
- Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1337–1345. doi: 10.1097/00004583-200111000-00015
- Green, S. B., and Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika* 74, 155–167. doi: 10.1007/s11336-008-9099-3
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med. Care* 44, S78–S94. doi: 10.1097/01.mlr.0000245454.12228.8f
- Hamilton, L. T. (2016). *Parenting to a degree: how family matters for college women's success*. Chicago: University of Chicago Press.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. doi: 10.1007/BF02289447
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., and Rosseel, Y. (2020). semTools: useful tools for structural equation modeling. R package version 0.5-3. Available online at: <https://CRAN.R-project.org/package=semTools> (Accessed February 16, 2026).
- Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociol. Methods Res.* 44, 486–507. doi: 10.1177/0049124114543236
- Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* 38, 52–54. doi: 10.5395/rde.2013.38.1.52
- Klasen, H., Woerner, W., Rothenberger, A., and Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu)--Übersicht und Bewertung erster Validierungs- und Normierungsbefunde [German version of the Strengths and Difficulties Questionnaire (SDQ-German)--overview and evaluation of initial validation and normative results]. *Prax. Kinderpsychol. Kinderpsychiatr.* 52, 491–502. doi: 10.23668/psycharchives.11726
- Kliem, S., Krieg, Y., Beller, J., Brähler, E., and Baier, D. (2021). Psychometric properties of the somatic symptom scale 8 (SSS-8) in a representative sample of German adolescents. *J. Psychosom. Res.* 149:110593. doi: 10.1016/j.jpsychores.2021.110593
- Kliem, S., Krieg, Y., Lohmann, A., and Mößle, T. (2020a). Evaluation of the universal prevention program Klasse2000 in fourth grade primary school children: protocol for a propensity score-matching approach. *JMIR Res. Protoc.* 9:e14371. doi: 10.2196/14371
- Kliem, S., Lohmann, A., Mößle, T., and Brähler, E. (2018). Psychometric properties and measurement invariance of the Beck hopelessness scale (BHS): results from a German representative population sample. *BMC Psychiatry* 18:110. doi: 10.1186/s12888-018-1646-6
- Kliem, S., Lohmann, A., Mößle, T., Foran, H. M., Hahlweg, K., Zenger, M., et al. (2019). Development and validation of a parenting scale short form (PS-8) in a representative population sample. *J. Child Fam. Stud.* 28, 30–41. doi: 10.1007/s10826-018-1257-3
- Kliem, S., Lohmann, A., Neumann, M., Glaubitz, C., Haselbach, S., Bergmann, M. C., et al. (2020b). Factor structure of the inventory of callous-unemotional traits in a representative sample of German 9th grade students. *J. Abnorm. Child Psychol.* 48, 43–55. doi: 10.1007/s10802-019-00590-x
- Kliem, S., Mößle, T., Rehbein, F., Hellmann, D. F., Zenger, M., and Brähler, E. (2015). A brief form of the perceived social support questionnaire (F-SozU) was developed, validated, and standardized. *J. Clin. Epidemiol.* 68, 551–562. doi: 10.1016/j.jclinepi.2014.11.003
- Kliem, S., Mößle, T., Zenger, M., and Brähler, E. (2014). Reliability and validity of the Beck depression inventory-fast screen for medical patients in the general German population. *J. Affect. Disord.* 156, 236–239. doi: 10.1016/j.jad.2013.11.024
- Kliem, S., Sachser, C., Lohmann, A., Baier, D., Brähler, E., Fegert, J. M., et al. (2025). Psychometric evaluation and community norms of the GAD-7, based on a representative German sample. *Front. Psychol.* 16:1526181. doi: 10.3389/fpsyg.2025.1526181
- Kliem, S., Sachser, C., Lohmann, A., Baier, D., Brähler, E., Gündel, H., et al. (2024). Psychometric evaluation and community norms of the PHQ-9, based on a representative German sample. *Front. Psych.* 15:1483782. doi: 10.3389/fpsyg.2024.1483782
- Komsta, L., and Novomeskay, F. (2015). Moments: moments, cumulants, skewness, kurtosis and related tests (R package version 0.14). Available online at: <https://CRAN.R-project.org/package=moments> (Accessed February 16, 2026).
- Kroenke, K., Spitzer, R. L., Williams, J. B., and Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics* 50, 613–621. doi: 10.1176/appi.psy.50.6.613
- Levy, P. (1967). The correction for spurious correlation in the evaluation of short-form tests. *J. Clin. Psychol.* 23, 84–86. doi: 10.1002/1097-4679(196701)23:1<84::AID-JCLP2270230123>3.0.CO;2-2
- Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol. Methods* 21, 369–387. doi: 10.1037/met0000093
- Lorber, M. F., Xu, S., Slep, A. M. S., Bulling, L., and O'Leary, S. G. (2014). A new look at the psychometrics of the parenting scale through the lens of item response theory. *J. Clin. Child Adolesc. Psychol.* 43, 613–626. doi: 10.1080/15374416.2014.900717
- Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., et al. (2010). A 4-item measure of depression and anxiety: validation and standardization of the patient health Questionnaire-4 (PHQ-4) in the general population. *J. Affect. Disord.* 122, 86–95. doi: 10.1016/j.jad.2009.06.019
- McLeod, J. D., and Shanahan, M. J. (1993). Poverty, parenting, and children's mental health. *Am. Sociol. Rev.* 58, 351–366. doi: 10.2307/2095905
- Milevsky, A., Schlechter, M., Netter, S., and Keehn, D. (2007). Maternal and paternal parenting styles in adolescents: associations with self-esteem, depression, and life satisfaction. *J. Child Fam. Stud.* 16, 39–47. doi: 10.1007/s10826-006-9066-9065
- Raboteg-Saric, Z., and Sakic, M. (2014). Relations of parenting styles and friendship quality to self-esteem, life satisfaction and happiness in adolescents. *Appl. Res. Qual. Life* 9, 749–765. doi: 10.1007/s11482-013-9268-0
- Rehbein, F., Kliem, S., Baier, D., Mößle, T., and Petry, N. M. (2015). Prevalence of internet gaming disorder in German adolescents: diagnostic contribution of the nine DSM-5 criteria in a state-wide representative sample. *Addiction* 110, 842–851. doi: 10.1111/add.12849
- Reitman, D., Currier, R. O., Hupp, S. D., Rhode, P. C., Murphy, M. A., and O'Callaghan, P. M. (2001). Psychometric characteristics of the parenting scale in a head start population. *J. Clin. Child Psychol.* 30, 514–524. doi: 10.1207/S15374424JCCP3004_08
- Rhoades, K. A., and O'Leary, S. G. (2007). Factor structure and validity of the parenting scale. *J. Clin. Child Adolesc. Psychol.* 36, 137–146. doi: 10.1080/15374410701274157
- Rosseel, Y. (2012). Llava: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. Available online at: <http://www.jstatsoft.org/v48/i02/> (Accessed February 16, 2026).
- Salari, R., Terreros, C., and Sarkadi, A. (2012). Parenting scale: which version should we use? *J. Psychopathol. Behav. Assess.* 34, 268–281. doi: 10.1007/s10862-012-9281-x
- Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res.* 8, 23–74. doi: 10.23668/psycharchives.12784

- Schmalbach, B., Irmer, J. P., and Schultze, M. (2019). ezCutoffs: fit measure cutoffs in SEM. R package version 1.0.2. Available online at: <https://CRAN.R-project.org/package=ezCutoffs> (Accessed February 16, 2026).
- Schultze, M. (2020). Stuart: subtests using algorithmic rummaging techniques. R package version 0.9.1. Available online at: <https://CRAN.R-project.org/package=stuart> (Accessed February 16, 2026).
- Steinberg, L., Elmen, J. D., and Mounts, N. S. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Dev.* 60, 1424–1436. doi: 10.1111/j.1467-8624.1989.tb04014.x
- Sulik, M. J., Blair, C., Mills-Koonce, R., Berry, D., and Greenberg, M. Family Life Project Investigators (2015). Early parenting and the development of externalizing behavior problems: longitudinal mediation through children's executive function. *Child Dev.* 86, 1588–1603. doi: 10.1111/cdev.12386
- Tran, T. D., Luchters, S., and Fisher, J. (2017). Early childhood development: impact of national human development, family poverty, parenting practices and access to early childhood education. *Child Care Health Dev.* 43, 415–426. doi: 10.1111/cch.12395
- Waters, L. (2015). Strength-based parenting and life satisfaction in teenagers. *Adv. Soc. Sci. Res. J.* 2, 158–173. doi: 10.14738/assrj.211.1551
- Wu, H., and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika* 81, 1014–1045. doi: 10.1007/s11336-016-9506-0
- Zhang, G., Jiang, G., Hattori, M., and Trichtinger, L. (2020). EFAutilities: utility functions for exploratory factor analysis. R package version 2.1.1. Available online at: <https://CRAN.R-project.org/package=EFAutilities> (Accessed February 16, 2026).