PRIFYSGOL
BANGOR
UNIVERSITY

Measuring dysfunctional parenting: Psychometrics of three versions of the Parenting Scale

Prandstetter, Katharina; Waller, Franziska; Heinrichs, Nina; Hutchings, Judy; Ward, Catherine ; Danila, Ingrid; Lachman, Jamie; Foran, Heather

Family Relations

[Cyswllt i'r cyhoeddiad / Link to publication](Link to publication)

19. Apr. 2024

RESEARCH

Measuring dysfunctional parenting: Psychometrics of three versions of the Parenting Scale

Katharina Prandstetter[1] | **Franziska Waller[1]** | **Nina Heinrichs[2]** |
Judy Hutchings[3] | **Catherine L. Ward[4]** | **Ingrid Dănilă[5]** |
Jamie M. Lachman[6] | **Heather M. Foran[1]**

[1]Institute of Psychology, University of Klagenfurt, Klagenfurt, Austria

[2]Department of Psychology, University of Bremen, Bremen, Germany

[3]Department of Psychology, Bangor University, Bangor, UK

[4]Department of Psychology, University of Cape Town, Cape Town, South Africa

[5]Department of Psychology, Babes-Bolyai University, Cluj-Napoca, Romania

[6]Department of Social Policy and Intervention, University of Oxford, Oxford, UK

Correspondence
Katharina Prandstetter, University of Klagenfurt, University Street 65-67, Klagenfurt 9020, Austria.
Email: katharina.prandstetter@aau.at

Abstract
Objective: This study assessed the psychometric properties of three versions of the Parenting Scale (PS; original PS, 13-item version, and 10-item version) in three European middle-income countries.
Background: The PS is one of the most frequently used questionnaires for measuring dysfunctional discipline strategies. Although its validity has been extensively investigated in American samples, there are mixed results regarding the recommended number of items and subscales, raising the question of replicability across European middle-income countries.
Method: Multigroup confirmatory factor analysis (MCFA) and item response theory (IRT) were applied to $N = 835$ parents from North Macedonia, Moldova, and Romania.
Results: All three versions were significantly correlated with parental- and child-related variables. Confirmatory factor analysis indicated the best model fit for the 10-item version, and configural and partial metric invariance across countries could be established for this version. Item response theory analyses also supported this measure.
Conclusions: Our findings show that the 10-item version performed better than the 13-item version and the original PS both overall and on the country level. Reliability values were somewhat lower than reported in studies from the United States.
Implications: The 10-item version constitutes a promising short measure for assessing dysfunctional parenting in European middle-income countries for researchers and practitioners.

Globally, 10% to 20% of young children and adolescents suffer from mental health problems (Kieling et al., 2011), with a disproportionately high number (approximately 85%) coming from low- and middle-income countries (LMICs; Erskine et al., 2017). As data from Global Burden of Disease (GBD) studies revealed, mental health disorders are among the leading noncommunicable diseases associated with reduced life expectancy and earlier deaths in adults and children (GBD 2019 Risk Factors Collaborators, 2020). A large body of research has highlighted the impact of parenting and other relationship-related behaviors on children's mental and physical health development (Pinquart, 2017; Stith et al., 2009). Dysfunctional discipline strategies featuring physical punishment, overt expression of anger, and permissive and lax behaviors have consistently been associated with the onset of behavioral and emotional problems in children, leading to internalizing and externalizing disorders (MacKenzie et al., 2015; Stormshak et al., 2000). Past research has also pointed to links between parental stress, dysfunctional parenting behaviors, and child behavioral problems (Crnic et al., 2005; Deater-Deckard, 1998; Neece et al., 2012), highlighting the importance of familial variables for children's mental health development. Children from socioeconomically disadvantaged families are particularly prone to mental health problems, being two to three times more likely to develop mental health disorders during childhood and adolescence (Reiss, 2013). Additionally, families in LMICs are more likely to be exposed to poverty, food insecurity, and violence, conditions that negatively impact parental and child mental health (Hillis et al., 2016) and also increase parenting challenges (Henninger & Luze, 2014; Kieling et al., 2011). Furthermore, parents exposed to poverty have a higher risk of using harsh and lax parenting strategies as well as physical punishment as means of disciplining the child (Bradley et al., 2001; Elder et al., 1995; Scaramella et al., 2008). In response to these findings, several prevention and intervention parenting programs based on principles such as the social learning theory (e.g., Triple P; Sanders, 2008; Parenting for Lifelong Health for Young Children [PLH-YC]; Jansen et al., 2021), attachment theory (e.g., Video-feedback intervention [VF]; Bakermans-Kranenburg et al., 2003; Circle of Security intervention [COS]; Hoffman et al., 2006) or a combination of both (Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline [VIPP-SD]; Juffer et al., 2017) have been developed. One major goal of these programs is to reduce child mental health problems by targeting dysfunctional parenting behaviors, using methods such as teaching positive parenting practices, setting limits, and promoting sensitive parenting. Overall, the effectiveness of parenting programs is well-documented, with empirical support demonstrating significant reductions in parental stress, dysfunctional parenting behavior, and children's internalizing and externalizing symptoms (Barlow & Coren, 2018; Gardner et al., 2010; Kaminski et al., 2008).

The use of valid and reliable measures to evaluate the efficacy and effectiveness of parenting programs is of utmost importance for several reasons. First, for many parenting interventions, significant changes in behavioral problems of children and dysfunctional parenting behaviors are used to determine their applicability and impact (Barlow & Coren, 2018; De Graaf et al., 2008; Dretzke et al., 2009). Second, psychometrically solid measures represent crucial tools for researchers and clinicians to screen families at risk and to offer support at early stages. Third, measurement tools are highly essential to enable researchers to empirically test theoretically established pathways validly and reliably. A review of the psychometric properties of peer-reviewed parenting measures, however, showed that the majority lacked sufficient psychometric support for use in professional and research practice (Hurley et al., 2014). Twenty of 25 questionnaires did not provide any information on norming, and approximately 44% did not include any scoring information. Only five measures showed evidence for acceptable psychometric properties, with the Parenting Scale (Arnold et al., 1993) being one of them.

THE PARENTING SCALE

The Parenting Scale (PS) was developed to assess dysfunctional parental discipline strategies (Arnold et al., 1993) and is one of the most frequently employed questionnaires in clinical and research practice on parenting (Pritchett et al., 2011). As previous research showed, several studies have utilized the PS to assess dysfunctional parenting as a main outcome of their intervention studies, given its sensitivity to change (De Graaf et al., 2008; Nowak & Heinrichs, 2008). As a 30-item self-report measure, it consists of three subscales: Laxness (highly permissive and inconsistent parenting behavior), Overreactivity (harsh, impulsive, and aggressive parenting), and Verbosity (repeated talking instead of taking action). Participants are presented with different situations (e.g., "When my child misbehaves …") and asked to indicate their parenting behavior on a Likert Scale ranging from 1 to 7 (e.g., $1 = I\ do\ something\ right\ away$ … to $7 = I\ do\ something\ later$), with higher scores reflecting higher levels of dysfunctional parenting.

In the last 2 decades, numerous studies have tried to replicate the original factor structure (Arnold et al., 1993), with relatively consistent empirical support for the use of Laxness and Overreactivity, but not Verbosity (Pritchett et al., 2011; Rhoades & O'Leary, 2007; Salari et al., 2012; Steele et al., 2005). Nevertheless, suggestions as to which and how many items best reflect the two remaining factors are varied (Irvine et al., 1999; Lorber et al., 2014; Reitman et al., 2001; Rhoades & O'Leary, 2007) and have initiated the development of several shorter versions. Reitman et al. (2001), for instance, did not find empirical support for the original three-factor structure and instead suggested the use of a modified Laxness and Overreactivity scale, each consisting of five items. In a later study, Rhoades and O'Leary (2007) successfully replicated the version recommended by Reitman et al. (2001) but proposed a 13-item measure with parental hostility as a new factor (this version of the PS is referred to as the "13-item version" in the paper). The Hostility scale was composed of items originally assigned to Overreactivity but which appeared to be more indicative of harsh parenting (i.e., physical punishment, verbal aggression). Another short version, which was recently discussed, is the one by Lorber et al. (2014; referred to as the "10-item version" in this paper), who investigated which items of the PS best discriminated parents with more serious parenting problems. This shorter form consisted of five lax and overreactive items each and included those items providing most information along the continuum of functional to dysfunctional discipline behaviors in mothers and fathers. Parenting Scale items for each version were very similar, with one Laxness item being different for the 10-item version for men compared to women (i.e., PS7 "Threaten things I know I won't do" for mothers and PS21 "Offer something nice to behave if 'no' fails" for fathers).

AIM OF THE CURRENT STUDY

Although the validity of the PS has been extensively investigated in high-income countries (HICs), with the majority of studies coming from the United States (Freeman & DeCourcey, 2007; Lorber et al., 2014; Reitman et al., 2001; Rhoades & O'Leary, 2007; Steele et al., 2005), information on its performance in European middle-income countries (MICs) is currently limited. Although some studies validated translated versions of the PS in their respective countries, such as Germany, Australia, and Sweden (Kliem et al., 2019; Naumann et al., 2010; Prinzie et al., 2007; Salari et al., 2012), none of these studies were conducted in MICs nor in former Communist-bloc countries, raising the question about the replicability of this measure in low-resource settings. As past research suggests, participants may benefit more from culturally adapted health interventions than those that have not undergone this process

(Griner & Smith, 2006; Sundell et al., 2016). For example, only recently, McWayne et al. (2017) suggested that positive parenting among low-income Black families may be different from positive parenting in other cultural contexts. To account for this, the authors developed the Black Parenting Strengths in Context measure to better honor the cultural diversity and context-specificity of low resource settings. This approach, however, is very cost intensive and limited in its applicability, given the large number of different cultures. Another, more economical way of investigating and acknowledging this aspect of measurement is to focus on instruments that have been widely accepted and to specifically test their psychometric properties across heterogeneous samples.

In line with this, the current study aims to address this research gap by utilizing an exploratory approach to examine the psychometrics of the PS in three middle-income European countries—North Macedonia, Republic of Moldova, and Romania. Following data from The World Bank (n.d.), 22%–25% of the population in these countries are living below the national poverty line and are affected by economic hardship. Furthermore, only a small number of support programs are available for families in these countries, while at the same time showing elevated levels of family difficulties and violence (Levav et al., 2004; UNICEF, 2014). For instance, a study by Lansford et al. (2020) conducted in 21 LMICs, including Moldova, showed that 30% to 40% of participants found violence toward the wife to be an adequate reaction in 1 out of 5 situations, and 55% of participants further reported that they had used physical violence as a means of disciplining their child in the last 30 days, highlighting the importance of taking into account different beliefs and attitudes present in different cultures. Drawing on open questions, such as the applicability of short versions of the PS and the need for psychometric investigations of these versions in European MICs, the following research questions were examined in the current study: (1) How well does the original PS perform compared to two short forms (10-item version and 13-item version) in terms of validity and reliability in a sample of $N = 835$ caregivers from three MICs? (2) Do the three versions of the PS perform equally well across North Macedonia, Republic of Moldova, and Romania with respect to validity and reliability?

METHOD

Participants and procedure

Data collection took place in North Macedonia, the Republic of Moldova, and Romania during March and April 2019 as part of a three-phase study (with the present data coming from the baseline assessment of Phase 2 [Lachman et al., 2019]). Potential participants were invited to take part in a parenting program for parents of children aged 2 to 9 years (Parenting for Lifelong Health for Young Children; PLH-YC) and were contacted by research coordinators via phone, letter, or in person. Participants were screened for eligibility and informed that they had the right to decline participation and/or withdraw from the program and/or the research evaluation at any time. Caregivers, aged 18 years or older, were eligible to participate if they had lived in the same household as their child for at least four nights a week in the previous month and reported that the child on whom they chose to focus on for the program showed elevated levels of child behavior problems based on scores of 10 or above on the Child and Adolescent Behavior Inventory oppositional defiant disorder subscale (eight items; Burns et al., 2015).

Data assessors were local research assistants with intensive training in ethics, informed consent, and interviewing techniques. Questionnaires were completed by each caregiver, either using an electronic tablet or a paper and pencil interview-assisted response sheet. For

their participation, all caregivers received a food/gift voucher for data completion at each assessment point. The study was approved by the Human Research Ethics Committee of the University of Klagenfurt and the Human Research Ethics Commission of the local institution of the respective country. Overall, 835 caregivers participated in the parenting program (96.0% female caregivers). The majority of caregivers reported being the biological parent of the child (92.1% biological mother, 3.8% biological father, 2.6% grandmother/grandfather, 1.3% other). Demographic information and sample characteristics can be found in Table 1.

Measures

Besides demographics, participants completed questionnaires on parenting behaviors, parental stress, and children's externalizing and internalizing symptoms.

TABLE 1   Demographic information of the sample

Sample characteristics	Sample (N = 835)		North Macedonia (n = 289)		Moldova (n = 284)		Romania (n = 262)	
	M (SD)	Median	M (SD)	Median	M (SD)	Median	M (SD)	Median
Caregiver (age)	36.3 (6.5)	36.0	37.6 (4.8)	37.0	34.8 (8.3)	33.0	36.4 (5.5)	36.0
Child (age)	5.7 (2.0)	6.0	6.0 (2.0)	6.0	5.4 (2.0)	5.0	5.6 (1.9)	5.0
	N	%	N	%	N	%	N	%
Child gender								
Boy	504	60.4	173	59.9	154	54.2	177	67.6
Girl	331	39.6	116	40.1	130	45.8	85	32.4
Education level								
No university/ college	221	25.3	50	17.3	120	42.3	41	15.6
University/ college	624	74.7	239	82.7	164	57.7	221	84.4
Literacy								
Cannot read/ reading difficulties	7	0.8	1	0.3	3	1.1	3	1.1
Little reading difficulties/ read easily	828	99.2	288	99.7	281	98.9	259	98.9
Marital status								
Single or widowed	34	4.1	16	5.5	12	4.2	6	2.3
In a relationship	46	5.5	10	3.5	21	7.4	15	5.7
Married	709	84.9	244	84.4	237	83.5	228	87.0
Separated/ divorced	46	5.5	19	6.6	14	4.9	13	5.0
Household income								
No income/other than salaries	59	7.1	10	3.5	39	13.7	10	3.8
Salaries and wages	776	92.9	279	96.5	245	86.3	252	96.2

Note: N = total sample size; n = subsample size for the respective countries.

Parenting Scale (PS)

The PS (Arnold et al., 1993) Laxness (11 items) and Overreactivity (10 items) subscales were used to measure dysfunctional parental discipline strategies. Participants were asked to rate their parenting behavior when encountering child misbehavior on a scale ranging from 1 to 7, with higher scores indicating more dysfunctional parenting strategies. For the purposes of this study, the original PS version (21 items) was administered, and the scores for the 13-item version and 10-item version were extracted from the data. Cronbach's α for the three versions can be found in the Results section.

Parenting of Young Children (PARYC)

Positive parenting was assessed using the PARYC (McEachern et al., 2012), a 21-item self-report measure in which parents were asked to indicate the amount of positive parenting behaviors on a scale from 1 = *never* to 7 = *always* (e.g., "Spend time with your child in ways that were fun for both of you?"). The PARYC includes three subscales: Positive Parenting, Monitoring, and Planning. Higher scores reflect more positive parenting. For the purposes of this study, only the total score was utilized (α = .89 for the total sample).

Parental Stress Scale (PSS)

Parental stress was measured utilizing the PSS (Berry & Jones, 1995). This self-report questionnaire comprises 18 items, which are rated on a Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree* (e.g., "I am happy in my role as a parent"). In addition to a total score, with higher values reflecting higher levels of parental stress, four subscales can be computed: Rewards, Stressors, Loss of control, and Satisfaction. In the current study, only the PSS total score was used (α = .80 for the total sample).

Child behavioral problems

Children's internalizing and externalizing behaviors were measured using the Child Behavior Checklist (CBCL) versions for younger (CBCL/1½–5 version) and older (CBCL/6–18 version) children (Achenbach, 1991; Achenbach & Rescorla, 2000). Parents were presented with 103 (younger children) or 113 items (older children) and asked to rate the frequency of certain child behaviors on a Likert scale (e.g., "Destroys things belonging to his/her family or other children"; 0 = *not true*; 1 = *somewhat or sometimes true*; 2 = *very true or often true*). In addition to a total score, scores for internalizing and externalizing were calculated. For the purposes of this study, a CBCL combined version (aggregated across older and younger children) was used (CBCL young: α = .94, CBCL old: α = .95 for the total sample). All questionnaires were translated into the local language (Macedonian, Moldovan, and Romanian) and translated back into English for verification of accuracy. The CBCL translations went through a translation process with the developers of the CBCL and bilingual child mental health experts in the local languages.

Analytical strategy

Overall, there were very low rates of missing data in this sample (<1%). To account for the non-normality of variables, maximum likelihood estimation with robust standard errors (MLR)

was used as estimator choice, and full information maximum likelihood (FIML) was utilized to account for missing data (Enders & Bandalos, 2001). The following statistical analyses were conducted to address the above-mentioned research questions: First, confirmatory factor analysis (CFA) was applied to investigate the validity of the different versions of the PS and obtain information on the performance of the PS in a middle-income European sample. Second, measurement invariance and chi-square difference testing (mean-adjusted Satorra-Bentler-scaled χ^2) were conducted to assess whether the factor structure and loadings of the PS versions were the same among the three countries. This approach seemed relevant because the equality of the psychometrics of the PS across three countries must be tested explicitly to acknowledge potential differences in the subsamples that might be overlooked when only analyzing the total sample. Furthermore, each country used a different translation of the measure in accordance with the main language in the respective country, highlighting the need for additional country-level analysis. Third, the generalized partial credit model (GPCM) was applied to obtain information on how well the PS items discriminate among parents with different levels of lax and overreactive parenting, as well as to identify potentially problematic items. This model was utilized to complement the CFA approach and enable a more nuanced item-level analysis of the three versions of the PS overall and on the country level. Furthermore, Pearson correlations were calculated to investigate convergent and divergent validity for the different PS versions, with respect to measures that have been theoretically linked to dysfunctional parenting behaviors, such as parental stress, positive parenting behaviors, and child mental health problems. To examine the reliability of the original PS, 13-item version, and 10-item version, Cronbach's α, including a 95% confidence interval (CI), was computed for the total sample and for each country.

Statistical analyses were conducted using Mplus (Version 8.4; Muthén & Muthén, 2019). The following fit indices were applied to evaluate the overall model fit: The standardized root-mean-square residual (SRMR), the root-mean-square error of approximation (RMSEA), the Tucker–Lewis index (TLI), and the comparative fit index (CFI). For SRMR and RMSEA, values <.05 are considered a good model fit. For the CFI and TLI, values >.90 indicate acceptable, and > .95 imply good model fit (Hu & Bentler, 1999).

RESULTS

Descriptive analysis

Analyses found significant differences in demographics among the three countries for caregiver age, $F(2, 531.40) = 12.49$, $p < .001$; child age, $F(2, 832) = 8.19$, $p < .001$; child gender, $F(2, 832) = 8.19$, $p < .001$; caregiver gender, $\chi^2(2) = 10.09$, $p = .006$; and caregiver relationship to the target child ($p < .001$, Fisher's exact test). Detailed information on demographic characteristics of the sample can be found in Table 1.

Intercorrelations

The total scores of the three versions were significantly intercorrelated. The Laxness scale of the 13-item version was not significantly correlated with its corresponding Overreactivity and Harshness scales, the original PS Overreactivity scale, and the Overreactivity scale of the 10-item version (see Table 2). All three versions were significantly associated with measures of positive parenting behaviors, parental stress, and child behavioral problems (see Table 2). However, correlations for the subscales differed considerably, with the Laxness subscale of the 13-item version displaying the lowest convergent and divergent correlations.

Item level analysis

Descriptive statistics of the items and item wordings can be found in Table 3. F tests were used to compare item means across the three countries accounting for family-wise error inflation. Item discrimination parameters and item location parameters from the GPCM for the total samples as well as separately for the three countries are presented in Table 4. For the interpretation of the item discrimination parameters, the following suggestions for cut-off values were used (Baker, 2001): negative values (indicative of problematic items and reverse response patterns), 0 (no discrimination), 0.01–0.34 (very low discrimination), 0.35–0.64 (low discrimination), 0.65–1.34 (moderate discrimination), 1.35–1.69 (high discrimination), and >1.70 (very high discrimination). Overall, item discrimination for the total sample and for the three countries ranged from low to moderate discrimination, except for PS12, which displayed a negative value for Moldova and Romania ($\alpha_j = -0.049$, $\alpha_j = -0.039$).

According to Baker (2001) and de Ayala (2008), values for item location parameters typically lie between a range of -3.00 to 3.00 with item location parameters less than -2.00 being classified as (very) easy, values around 0.00 as average, and values greater than 2.00 classified as (very) difficult, depending on the type of scale and context of assessment. Parenting Scale items ranged from easy to (very) difficult, with notable differences for PS12 across countries, as the item location parameters for the total sample and North Macedonia can be classified as relatively difficult ($b_j = 2.556$, $b_j = 1.472$), while for Moldova and Romania the parameters can be classified as average to easy ($b_j = -0.190$; $b_j = -0.652$). For a more nuanced item-level analysis, item information curves (IIC) were plotted for Laxness and Overreactivity items, which represent a helpful tool for identifying the amount of information each item provides along the continuum of functional to dysfunctional parenting (see Figure 1 and Figure 2). Figure 1 indicates that for Laxness, PS30 provided most information, particularly from the mid- to higher range of Laxness. Items PS19 and PS20 provided most information in the mid-range of Laxness, and overall, PS12 and PS7 provided least information along the whole continuum of lax parenting behaviors. Figure 2 displays the respective IICs for Overreactivity and shows that PS28 provides most information at the higher level and PS10 at the lower to mid-level of Overreactivity. Item PS9 was least informative along the continuum of low to high levels of overreactive parenting.

Factorial validity

Parenting scale original version

Model fit for the total sample was unsatisfactory (CFI = .757; TLI = .728; RMSEA = .051). While PS12 showed the lowest factor loading for Laxness, PS30 displayed the highest factor loading (see Table 5). Overall, loadings for Overreactivity ranged from .181 to .647. When performing the CFA separately for each country, results revealed that PS12 showed a negative factor loading in Moldova and Romania but not North Macedonia ($-.165$, $-.120$, and $.369$, respectively). Overall, the model fit was unsatisfactory, with slightly higher values for North Macedonia (North Macedonia: CFI = .819; TLI = .798; RMSEA = .051; Moldova: CFI = .767; TLI = .740; RMSEA = .045; Romania: CFI = .700; TLI = .665; RMSEA = .064).

10-item version

The 10-item version yielded excellent model fit for the total sample (CFI = .938; TLI = .918; RMSEA = .041), as well as within the different countries (North Macedonia: CFI = .954;

TABLE 2 Correlations for construct validity

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1 Original PS	1													
2 PS 10-item version	.883	1												
3 PS 13-item version	.870	.744	1											
4 Original LAX	.785	.683	.536	1										
5 Original OR	.713	.641	.787	.126	1									
6 LAX 10-item version	.707	.815	.454	.851	.170	1								
7 OR 10-item version	.668	.741	.728	.166	.883	.215	1							
8 LAX 13-item version	.587	.481	.530	.832	-.002	.664	.004	1						
9 OR 13-item version	.659	.635	.783	.128	.910	.197	.841	-.004	1					
10 HARSH 13-item version	.405	.302	.524	.106	.529	.075	.422	.053	.302	1				
11 PARYC	-.424	-.445	-.393	-.317	-.320	-.341	-.355	-.216	-.347	-.176	1			
12 PSS	.373	.371	.361	.231	.337	.258	.327	.155	.348	.168	-.463	1		
13 CBCL Extern.	.281	.315	.341	.139	.292	.134	.376	.144	.297	.181	-.259	.374	1	
14 CBCL Intern.	.184	.237	.223	.098	.184	.125	.253	.098	.216	.082	-.235	.310	.642	1

Note: CBCL Extern. = CBCL Externalizing Problems; CBCL Intern. = CBCL Internalizing Problems; HARSH = Harshness; LAX = Laxness; OR = Overreactivity; PARYC = Positive Parenting; PS = Parenting Scale; PSS = Parental Stress Scale. All correlations are significant at $p < .001$ for $r \geq .90$ or $r \leq -0.90$ and at $p < .05$ for $r \geq 0.60$ or $r \leq -0.60$.

TABLE 3 Descriptive statistics of the Parenting Scale and country differences

Laxness	Total sample		North Macedonia		Moldova		Romania	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
PS7*: "I threaten to do things that … I'm sure I can carry out – I know I won't actually do"	4.09	2.34	4.42	2.27	3.76	2.37	4.07	2.35
PS8**: "I am the kind of parent that … Sets limits on what my child is allowed to do – Lets my child do whatever he/she wants"	2.34	1.66	2.73	1.63	1.94	1.70	2.36	1.54
PS12**: "When I want my child to stop doing something … I firmly tell my child to stop – I coax or beg my child to stop"	3.51	2.26	2.70	2.13	3.93	2.24	3.96	2.17
PS15: "When we're not at home … I handle my child the way I do at home – I let my child get away with a lot more"	2.61	2.02	2.43	1.90	2.74	2.29	2.67	1.82
PS16**: "When my child does something I don't like … I do something about it every time it happens – I often let it go"	1.96	1.45	2.21	1.56	1.89	1.58	1.76	1.09
PS19**: "When my child doesn't do what I ask … I often let it go or end up doing it myself – I take some other action"	3.65	2.37	3.05	2.11	4.35	2.64	3.55	2.14
PS20**: "When I give a fair threat or warning … I often don't carry it out – I always do what I said"	3.82	2.14	3.62	1.95	4.58	2.35	3.23	1.86
PS21**: "If saying 'no' doesn't work … I take some other kind of action – I offer my child something nice so he/she will behave"	3.07	2.19	2.74	2.00	3.65	2.56	2.81	1.83
PS24**: "If my child misbehaves and then acts sorry … I handle the problem like I usually would – I let it go that time"	3.65	2.46	3.16	2.13	4.01	2.78	3.79	2.36
PS26**: "When I say my child can't do something … I let my child do it anyway – I stick to what I said"	3.30	2.18	3.30	2.03	4.02	2.52	2.53	1.60
PS30**: "If my child gets upset when I say 'no' … I back down and give in to my child – I stick to what I said"	2.79	1.96	2.46	1.66	3.36	2.39	2.54	1.58
PS3: "When I'm upset or under stress … I am picky and on my child's back – I am not more picky than usual"	3.44	2.24	3.51	2.15	3.37	2.47	3.45	2.06

(Continues)

TABLE 3 (Continued)

Laxness	Total sample		North Macedonia		Moldova		Romania	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
PS6**: "When my child misbehaves … I usually get into a long argument with my child – I don't get into an argument"	4.70	2.10	4.29	2.02	5.20	2.24	4.61	1.94
PS9**: "When my child misbehaves … I give my child a long lecture – I keep my talks short and to the point"	3.40	2.34	3.85	2.35	2.93	2.43	3.42	2.13
PS10: "When my child misbehaves … I raise my voice or yell – I speak to my child calmly"	4.22	2.03	4.21	2.01	4.31	2.17	4.12	1.88
PS14**: "After there's been a problem with my child … I often hold a grudge – Things get back to normal quickly"	2.32	2.00	3.28	2.41	1.63	1.53	2.01	1.48
PS17**: "When there is a problem with my child … Things build up and I do things I don't mean to do – Things don't get out of hand"	2.92	1.97	2.42	1.69	3.05	2.22	3.33	1.83
PS18**: "When my child misbehaves I spank, slap, grab, or hit my child … Never or rarely – Most of the time"	1.76	1.34	1.52	1.06	1.94	1.62	1.82	1.26
PS22*: "When my child misbehaves … I handle it without getting upset – I get so frustrated or angry that my child can see I'm upset"	4.17	2.21	3.99	2.09	4.48	2.46	4.03	2.01
PS25: "When my child misbehaves … I rarely use bad language or curse – I almost always use bad language"	1.51	1.22	1.56	1.20	1.38	1.23	1.62	1.24
PS28: "When my child does something I don't like, I insult my child, say mean things, or call my child names … Never or rarely – Most of the time"	1.34	0.98	1.24	0.73	1.39	1.14	1.39	1.01

Note: F tests were used to compare item means across the three countries accounting for family-wise error inflation.
*Statistically significant group differences, $p < .05$.
**Statistically significant group differences, $p < .001$.

TLI = .939; RMSEA = .038; Moldova: CFI = .946; TLI = .928; RMSEA = .033; Romania: CFI = .918; TLI = .891; RMSEA = .053). For the total sample, loadings ranged from .263 to .592 for Laxness, and .284 to .627 for Overreactivity (see Table 6 for more information).

TABLE 4 Item discrimination parameters for the total sample and three different countries for lax and overreactive parenting

	Total sample		North Macedonia		Moldova		Romania	
	λj	bj	λj	bj	λj	bj	λj	bj
Laxness								
PS7	0.131	−0.146	0.319	−0.389	0.090	0.426	0.133	−0.079
PS8	0.243	2.148	0.473	1.078	0.256	2.223	0.423	1.906
PS12	0.039	2.556	0.201	1.472	**−0.049**	−0.190	**−0.039**	−0.652
PS15	0.149	2.110	0.165	2.273	0.162	1.402	0.120	3.814
PS16	0.250	2.605	0.340	1.743	0.317	2.218	0.228	4.758
PS19	0.318	0.300	0.267	0.989	0.267	−0.199	0.336	0.456
PS20	0.380	0.041	0.519	0.227	0.192	−0.584	0.600	0.469
PS21	0.177	1.078	0.229	1.382	0.087	0.566	0.218	1.769
PS24	0.134	0.521	0.160	1.296	0.121	0.011	0.122	0.524
PS26	0.269	0.559	0.161	0.819	0.188	0.008	0.848	1.089
PS30	0.583	0.826	0.661	1.227	0.408	0.468	0.749	1.158
Overreactivity								
PS3	0.273	0.697	0.280	0.705	0.233	0.613	0.341	0.838
PS6	0.216	−0.620	0.239	0.002	0.136	−1.519	0.326	−0.418
PS9	0.102	1.354	0.086	0.545	0.130	1.533	0.148	1.161
PS10	0.530	−0.111	0.455	−0.060	0.517	−0.210	0.641	−0.014
PS14	0.187	2.096	0.349	0.585	0.317	2.194	0.138	5.003
PS17	0.358	1.047	0.532	1.443	0.279	0.932	0.456	0.682
PS18	0.348	2.417	0.604	2.451	0.264	2.231	0.314	3.268
PS22	0.404	−0.130	0.576	0.059	0.262	−0.385	0.524	−0.058
PS25	0.280	2.982	0.369	2.509	0.219	3.236	0.230	4.332
PS28	0.433	2.790	0.822	2.325	0.489	2.282	0.175	5.575

Note: λj = Item discrimination parameters; bj = Item locations parameters. Following Baker (2002), items' discriminatory power is characterized as follows: no (0), very low (0.01–0.34), low (0.35–0.64), moderate (0.65–1.34) and high (1.35–1.69). Negative item discrimination parameter values are indicative of problematic items and are marked bold.

13-item version

Overall, this version did not fit the data well (total sample: CFI = .781; TLI = .725; RMSEA = .056). However, fit information varied considerably across the three countries, with North Macedonia showing the best fit indices (CFI = .929; TLI = .911; RMSEA = .037), and Moldova (CFI = .741; TLI = .675; RMSEA = .063) and Romania (CFI = .736; TLI = .668; RMSEA = .075) displaying unsatisfactory model fit. For the total sample, loadings ranged from .066 to .670 for Laxness and from .313 to .644 for Overreactivity. Loadings for Harshness ranged from .474 to .532 for the total sample (see Table 6 for more results). Confirmatory factor analysis results on the country level indicated a negative factor loading of −.187 for Moldova and − .587 for Romania; however, convergence problems emerged when trying to fit the three-factor version to the Romanian sample. After checking for inter-item correlations of the respective Laxness and Overreactivity scales, item PS12 was identified as problematic, indicating significant negative correlations with items from the Overreactivity scale (ranging from −.082 to −.535) and nonsignificant correlations with items from the same scale (ranging from −.077 to −.109).
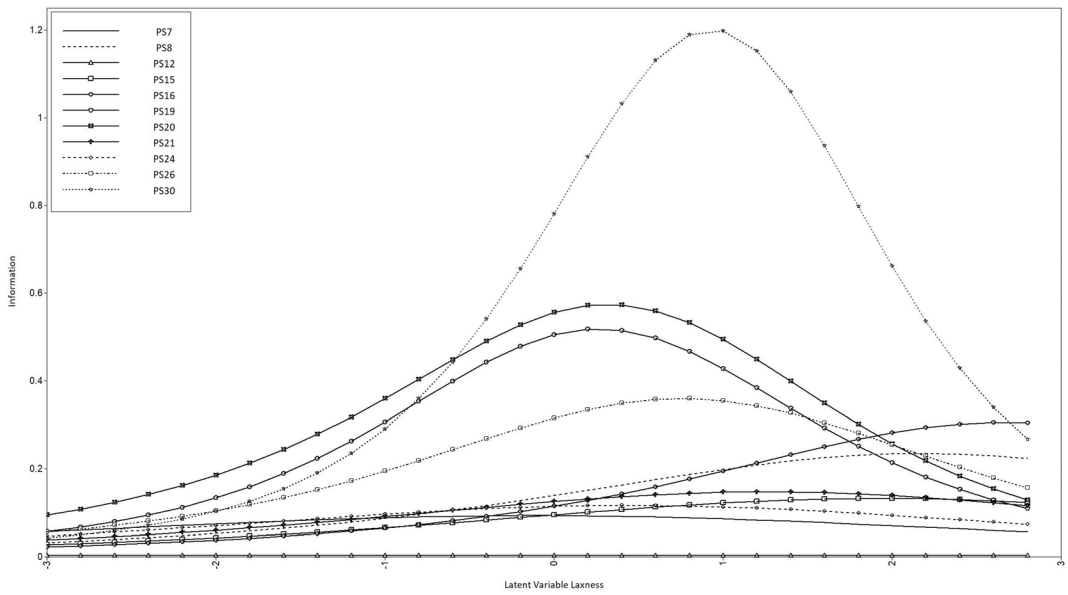
FIGURE 1 Item information curves (IIC) of the Laxness Scale *Note*. Analyses were run across all Laxness items utilizing the total sample of $N = 835$. For wording of the respective items, see Table 3.
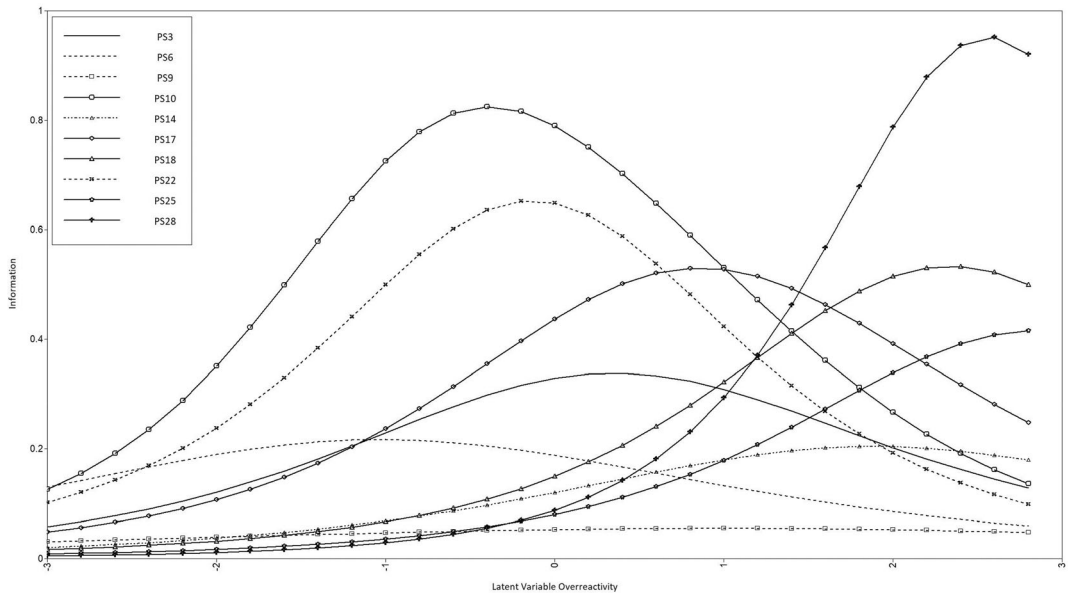
FIGURE 2 Item information curves (IIC) of the Overreactivity Scale *Note*. Analyses were run across all Overreactivity items utilizing the total sample of $N = 835$. For wording of the respective items, see Table 3.

Factorial invariance across countries

Tests for measurement invariance across countries were only conducted for the 10-item version, given the unsatisfactory model fit for the original PS Laxness and Overreactivity scales and the 13-item version. Results showed that configural invariance was supported for the 10-item

TABLE 5 CFA loadings and internal consistencies for the original PS version

Original version	Total sample (N = 835)	Moldova (n = 284)	North Macedonia (n = 289)	Romania (n = 262)
Laxness				
PS7	.270	.197	.497	.271
PS8	.343	.353	.540	.449
PS12	.064	−.165	.369	−.120
PS15	.288	.356	.295	.192
PS16	.322	.413	.422	.209
PS19	.535	.498	.462	.509
PS20	.547	.361	.598	.630
PS21	.323	.185	.372	.311
PS24	.304	.293	.316	.268
PS26	.459	.414	.294	.655
PS30	.628	.592	.624	.623
Cronbach's α	.632	.560	.710	.600
95% CI	[.593, .668]	[.480, .632]	[.658, .757]	[.525, .668]
Overreactivity				
PS3	.472	.485	.453	.510
PS6	.386	.271	.379	.483
PS9	.218	.286	.181	.286
PS10	.632	.642	.569	.664
PS14	.325	.386	.582	.202
PS17	.526	.507	.599	.575
PS18	.370	.337	.464	.328
PS22	.579	.470	.647	.611
PS25	.286	.227	.363	.265
PS28	.327	.366	.395	.196
Cronbach's α	.667	.644	.714	.681
95% CI	[.632, .700]	[.580, .703]	[.663, .761]	[.620, .735]

Note: CFA = confirmatory factor analysis; CI = confidence interval; PS = Parenting Scale.

measure: $\chi^2(102) = 152.704$, $p < .001$, RMSEA = .042 (.027–.056), CFI = .938, TLI = .918. Metric invariance, $\chi^2(16) = 39.283$, $p < .001$, and scalar invariance, $\chi^2(32) = 188.344$, $p < .001$, could not be established for the 10-item version.

Subsequent pairwise country comparisons showed that metric invariance was supported for North Macedonia versus Romania, $\chi^2(8) = 14.083$, $p = .080$, but not for North Macedonia versus Moldova, $\chi^2(8) = 37.147$, $p < .001$, and Moldova versus Romania, $\chi^2(8) = 16.163$, $p = .040$. To establish partial metric invariance, non-invariant items were identified for the respective countries. These included PS7 and PS20 for North Macedonia versus Moldova and PS20 for Moldova versus Romania. After allowing these items to be freely estimated, partial metric invariance could be established for the 10-item version: $\chi^2(12) = 13.559$, $p = .310$; North Macedonia versus Moldova: $\chi^2(6) = 7.613$, $p = .268$, and Moldova versus Romania: $\chi^2(7) = 10.332$, $p = .171$.

TABLE 6 CFA loadings and internal consistencies for the 10-item version and 13-item version

	Total sample (N = 835)	Moldova (n = 284)	North Macedonia (n = 289)	Romania (n = 262)
10-item version: Laxness				
PS7	.263	.197	.272	.271
PS19	.543	.498	.526	.509
PS20	.590	.361	.610	.630
PS26	.498	.414	.467	.655
PS30	.592	.592	.641	.623
Cronbach's α	.605	.539	.628	.635
95% CI	[.561, .646]	[.449, .619]	[.556, .692]	[.561, .701]
10-item version: Overreactivity				
PS6	.407	.271	.414	.483
PS10	.627	.642	.611	.664
PS17	.490	.507	.595	.575
PS22	.613	.470	.619	.611
PS28	.284	.366	.408	.196
Cronbach's α	.602	.543	.631	.634
95% CI	[.558, .643]	[.454, .622]	[.559, .694]	[.559, .699]
13-item version: Laxness				
PS12	.263	−.187	.359	−.587
PS16	.543	.288	.407	.189
PS19	.590	.504	.480	.152
PS21	.498	.162	.377	−.098
PS30	.592	.690	.642	.129
Cronbach's α	.443	.338	.539	.336
95% CI	[.381, .501]	[.207, .452]	[.449, .618]	[.199, .455]
13-item version: Overreactivity				
PS3	.407	.498	.445	.480
PS6	.627	.278	.366	.481
PS10	.490	.662	.550	.750
PS14	.613	.387	.585	.176
PS17	.284	.490	.598	.483
Cronbach's α	.581	.571	.637	.611
95% CI	[.535, .624]	[.487, .645]	[.566, .699]	[.531, .681]
Hostility				
PS18	.474	.423	.611	.455
PS25	.479	.399	.495	.641
PS28	.532	.583	.595	.414
Cronbach's α	.477	.426	.558	.485
95% CI	[.412, .536]	[.300, .533]	[.462, .640]	[.367, .585]

Note: CFA = confirmatory factor analysis; CI = confidence interval; PS = Parenting Scale.

Reliability analysis

PS original version

The original PS version yielded the following reliability values: Total score $\alpha = .663$ [.629, .695]; Laxness $\alpha = .632$ [.593, .668]; Overreactivity $\alpha = .667$ [.632, .700]. Results differed for the three different countries, with North Macedonia showing the highest values for internal consistency (total score $\alpha = .737$ [.691, .779]; Laxness $\alpha = .710$ [.658, .757]; Overreactivity $\alpha = .714$ [.663, .761]) and Moldova and Romania displaying values below $\alpha = .700$ (Moldova: total score $\alpha = .607$ [.538, .671]; Laxness $\alpha = .560$ [.480, .632]; Overreactivity $\alpha = .644$ [.580, .703]; Romania: total score $\alpha = .646$ [.581, .705]; Laxness $\alpha = .600$ [.525, .668]; Overreactivity $\alpha = .681$ [.620, .735]).

10-item version

Internal consistency for the 10-item version was as follows: $\alpha = .638$ [.600, .673] for the total score; $\alpha = .605$ [.561, .646] for Laxness; $\alpha = .602$ [.558, .643] for Overreactivity. Again, reliability indices differed for the three countries with Romania showing the highest values (total score $\alpha = .676$ [.614, .732]; Laxness $\alpha = .635$ [.561, .701]; Overreactivity $\alpha = .634$ [.559, .699]). For North Macedonia and Moldova, values ranging from $\alpha = .539$ to .644 (North Macedonia total score $\alpha = .644$ [.579, .702]; Laxness $\alpha = .628$ [.556, .692]; Overreactivity $\alpha = .631$ [.559, .694]; Moldova total score $\alpha = .569$ [.490, .640]; Laxness $\alpha = .539$ [.449, .619]; Overreactivity $\alpha = .543$ [.454, .662]) were obtained.

13-item version

Overall, the 13-item version yielded the lowest reliability values of $\alpha = .518$ [.469, .565] for the total score; $\alpha = .443$ [.381, .501] for Laxness, $\alpha = .581$ [.535, .624] for Overreactivity; and $\alpha = .477$ [.412, .536] for Harshness. On country level, values for North Macedonia were the highest, with $\alpha = .635$ [.570, .694] for the total score; $\alpha = .539$ [.449, .618] for Laxness; $\alpha = .637$ [.566, .699] for Overreactivity, and $\alpha = .558$ [.462, .640] for Harshness. For Moldova and Romania values ranged from $\alpha = .336$ to .611 (Moldova: total score $\alpha = .445$ [.346, .536]; Laxness $\alpha = .338$ [.207, .452]; Overreactivity $\alpha = .571$ [.487, .645]; Harshness $\alpha = .426$ [.300, .533]; Romania: total score $\alpha = .437$ [.332, .533]; Laxness $\alpha = .336$ [.199, .455]; Overreactivity $\alpha = .611$ [.531, .681]; Harshness $\alpha = .485$ [.367, .585]).

DISCUSSION

The PS is one of the most commonly used questionnaires for assessing dysfunctional parenting practices in young children (Pritchett et al., 2011). Due to its sensitivity to intervention impact, it has been frequently employed in clinical as well as research practice (Sanders, 2008; Thomas & Zimmer-Gembeck, 2007). Although the psychometrics of the PS have been extensively investigated in HICs, such as the United States, the literature on its performance in European MICs is sparse. Furthermore, the majority of past studies focused on exploratory factor analysis (EFA) and/or CFA approaches and have mostly analyzed differences in the PS's performance across child and parental age, child gender, parental gender, or parental education levels (Karazsia et al., 2008; Kliem et al., 2019; Steele et al., 2005). However, to date, it has not yet been examined how well different versions of the PS perform across European MICs. This paper,

therefore, aimed to address this research gap by investigating the psychometrics of the PS across North Macedonia, Moldova, and Romania.

Overall, the 10-item version showed the best model fit for the total sample as well as for the three countries separately. While configural invariance was supported for this measure, metric invariance could not be established, given that strengths of the factor loadings varied across the three countries. However, partial metric invariance could be established after removing two constraints for PS7 and PS20 for Moldova versus Romania and PS7 for North Macedonia versus Moldova. Furthermore, IICs indicated that items from the 10-item measure—such as PS30, PS19, and PS20 for Laxness and PS28, PS22, and PS10 for Overreactivity—were among the most informative items from the PS in the current sample. The poorer performance of the original measure, as well as the 13-item version compared to the 10-item short form, could be attributed to problematic items, such as PS12, which were not included in the 10-item version and thus could have reduced misfit in the CFA model. However, although PS12 worked for North Macedonia in the original PS and the 13-item version, measurement-related issues became apparent in Moldova and Romania, displaying problematic factor loadings and negative item discrimination parameters in these two countries. Further country-level analysis indicated that compared to Moldova (CBCL Extern.: 45%; CBCL Intern.: 53%) and Romania (CBCL Extern.: 37%; CBCL Intern.: 28%), parents in North Macedonia reported lower levels of child behavioral and emotional problems (CBCL Extern.: 18%; CBCL Intern.: 19%; calculated on t values ≥ 60), with values comparable to the validation samples of the original PS and the 13-item version (Arnold et al., 1993; Rhoades & O'Leary, 2007). These results point to potentially worse applicability of the 13-item version to countries with elevated levels of child behavioral and emotional problems. Additionally, Laxness IICs indicated that PS12 provided very little information along the continuum from low to high levels of dysfunctional parenting. These results suggest problems related to the wording and content of this item, pointing to the difficulty of adapting it to a different cultural context than that for which it was originally developed. Indeed, consultation with the Moldovan and Romanian research teams indicated that while the item content might be equally captured in both languages, problems could be related to a potentially less discriminative response format of PS12 in Moldovan and Romanian compared to the original English wording ("I firmly tell my child to stop … I coax or beg my child to stop"). This is also in line with research suggesting that concepts of parental discipline strategies may be perceived differently across cultures and social backgrounds (Bozicevic et al., 2016; Chao, 2000; McWayne et al., 2017), highlighting the relevance of cost-effective cultural adaptation processes of parenting measures, such as the PS.

Taken together, all three versions investigated in this paper were significantly associated with measures of positive parenting, parental stress, and child behavioral and emotional problems, which is in line with results from studies evaluating the validity of the PS in other European or American samples. For instance, Rhoades and O'Leary (2007) investigated the psychometrics of the 13-item version in 453 couples from New York and found that Overreactivity, Laxness, and Hostility were significantly associated with child externalizing ($r = .33$, $r = .15$, and $r = .27$, respectively), and impulsive behaviors ($r = .26$, $r = .14$, $r = .21$) at the $p < .01$ or $p < .001$ level. In the study by Lorber et al. (2014), the authors analyzed two American subsamples ($N = 453$, and $N = 399$), showing that the original PS, the 13-item version, and the 10-item version were significantly related to child externalizing behaviors for both fathers and mothers at the $p < .05$ level ($r_{\text{mothers}} = .47$, $r_{\text{fathers}} = .39$; $r_{\text{mothers}} = .42$, $r_{\text{fathers}} = .32$. $r_{\text{mothers}} = .45$, $r_{\text{fathers}} = .34$, respectively). Salari et al. (2012) recruited 617 mothers and 430 fathers from Sweden and also found significant associations for the original PS version and the 13-item version with child emotional ($r = .16$, $r = .16$), and behavioral problems ($r = .45$, $r = .42$), as well as parental stress ($r = .29$, $r = .25$) at the $p < .05$ level. Results of the current study, therefore, align with findings from parents in the United States and other European samples, indicating that associations of the three versions of the PS with measures on parental stress

and child emotional and behavioral problems are somewhat replicable across different countries and cultures. While not all studies, including the present one, utilized the same measure to assess convergent and divergent validity (for example, Rhoades & O'Leary [2007] used the Eyberg Child Behavior Inventory, while Lorber et al. [2014] applied the McArthur Health and Behavior Questionnaire, and Salari et al. [2012] utilized the Strengths and Difficulty Questionnaire as well as the Eyberg Child Behavior Inventory), these findings provide preliminary evidence of convergent and divergent validity of the PS as a measure for assessing dysfunctional parenting behaviors in various samples.

The psychometric results of the current study were most supportive of the 10-item version, given the good model fit overall and on the country level, results of the IICs, and configural and partial metric invariance across North Macedonia, Moldova, and Romania. Although reliability of the 10-item version was lower than what was reported in the original study, reliability indices of this version were similar to the corresponding long version in the current sample (10-item version: $\alpha = .638$; original PS: $\alpha = .663$). As Lorber and Slep (2018) showed, internal consistency is not always a solid indicator of the measurement precision of a questionnaire. Although the authors received low values of internal consistency for the Parent–Child Conflict Tactics Scale (CTS-PC) in their study, further IRT analysis showed that these items provided important evidence for differentiating parents with serious parenting problems from those without. Following these results, lower reliability coefficients in the current sample do not necessarily indicate that the PS is an unprecise measure for assessing dysfunctional parenting in the current sample. Rather, low internal consistency could be attributed to heterogeneity in discipline behavior: For instance, some parents might report holding a grudge after child misbehavior but might not yell at their children or call them mean names when encountering a difficult situation. Such heterogeneity in parenting strategies can potentially result in lower correlations between items assigned to the same subscale, which also influences reliability indices of this measure.

Implications for research and practice

Overall, researchers and practitioners are advised to use the 10-item version to evaluate dysfunctional parenting behaviors in European samples from MICs, given its more economical applicability (10 instead of 21 items) and better psychometric performance compared to the original PS and the 13-item version. Nevertheless, future investigations of the 10-item version might be beneficial to further improve the psychometrics of this measure. For example, PS7 displayed low factor loadings and item discrimination parameters in this sample and was identified as non-invariant in multigroup analyses, indicating room for improvement, such as the revision or replacement of this item. Moreover, we recommend future revisions of the PS not solely be based on psychometric analyses but also include theoretical considerations on dysfunctional parenting. In particular, the importance of cultural aspects and different parenting norms, as well as concepts, needs to be taken more into consideration to ensure a valid assessment of dysfunctional parenting across various samples.

Strengths and limitations

Recent calls for the implementation of evidence-based and cost-effective interventions in LMICs have also raised awareness of the need to investigate the cross-cultural and cross-country adaptability of instruments used in prevention and intervention programs. This paper is the first to evaluate the psychometric properties of three versions of the PS in a sample of three European MICs, providing researchers and clinicians with recommendations on the future

use of this questionnaire in low-resource settings. Although the participation rate was relatively high (83.1%), the majority of parents taking part in the program were female, and therefore, no conclusions can be made about the performance of the different PS versions among male caregivers. Recruitment of fathers for family research has remained a challenging topic (Fabiano, 2007; Phares et al., 2005) and certainly requires more attention in future research. One significant strength of this study, however, lies in the sample, offering insight into the applicability and performance of the PS in three under-researched European MICs. Future research, however, is needed to investigate the generalizability of the current results to other European low-resource MICs. As expected, means for dysfunctional parenting behaviors reported in the current sample (3.08, $SD = 0.72$) were somewhat higher compared to American studies (2.68, $SD = 0.85$); Lorber et al., 2014), underlining the importance of further research and interventions efforts in these countries, as well as the need to address measurement issues and prevention efforts for child and parental mental health simultaneously.

ORCID

Katharina Prandstetter https://orcid.org/0000-0002-6476-8284
Franziska Waller https://orcid.org/0000-0002-4576-4325
Nina Heinrichs https://orcid.org/0000-0002-8301-5798
Judy Hutchings https://orcid.org/0000-0003-2562-8104
Catherine L. Ward https://orcid.org/0000-0001-8727-4175
Ingrid Dănilă https://orcid.org/0000-0003-4190-8739
Jamie M. Lachman https://orcid.org/0000-0001-9475-9218
Heather M. Foran https://orcid.org/0000-0001-6659-5500

REFERENCES

Achenbach, T. M. (1991). *Manual for the youth self-report and 1991 profile*. University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool form & profiles*. University of Vermont, Department of Psychiatry.

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment*, *5*(2), 137–144. https://doi.org/10.1037/1040-3590.5.2.137.

Baker, F. B. (2001). *The basics of Item Response Theory* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. *Psychological Bulletin*, *129*(2), 195–215. https://doi.org/10.1037/0033-2909.129.2.195.

Barlow, J., & Coren, E. (2018). The effectiveness of parenting programs: A review of Campbell Reviews. *Research on Social Work Practice*, *28*(1), 99–102. https://doi.org/10.1177/1049731517725184.

Berry, J. O., & Jones, W. H. (1995). The Parental Stress Scale: Initial psychometric evidence. *Journal of Social and Personal Relationships*, *12*(3), 463–472. https://doi.org/10.1177/0265407595123009.

Bozicevic, L., De Pascalis, L., Schuitmaker, N., Tomlinson, M., Cooper, P. J., & Murray, L. (2016). Longitudinal association between child emotion regulation and aggression, and the role of parenting: A comparison of three cultures. *Psychopathology*, *49*(4), 228–235. https://doi.org/10.1159/000447747.

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*(6), 1844–1867. https://doi.org/10.1111/1467-8624.t01-1-00382.

Burns, G., Lee, S., Servera, M., McBurnett, K., & Becker, S. P. (2015). *Child and Adolescent Behavior Inventory–Parent Version* 1.0.

Chao, R. K. (2000). The parenting of immigrant Chinese and European American mothers: Relations between parenting styles, socialization goals, and parental practices. *Journal of Applied Developmental Psychology*, *21*(2), 233–248. https://doi.org/10.1016/S0193-3973(99)00037-4.

Crnic, K., Gaze, C., & Hoffman, C. (2005). Cumulative parenting stress across the preschool period: Relations to maternal parenting and child behaviour at age 5. *Infant and Child Development*, *14*(2), 117–132. https://doi.org/10.1002/icd.384.

de Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. Guilford Press.

De Graaf, I., Speetjens, P., Smit, F., De Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple P positive parenting program on parenting: A meta-analysis. *Family Relations*, 57(5), 553–566. https://doi.org/10.1111/j.1741-3729.2008.00522.x.

Deater-Deckard, K. (1998). Parenting stress and child adjustment: Some old hypotheses and new questions. *Clinical Psychology: Science and Practice*, 5(3), 314–332. https://doi.org/10.1111/j.1468-2850.1998.tb00152.x.

Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., Taylor, R. S., Sandercock, J., & Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct problems: A systematic review of randomised controlled trials. *Child and Adolescent Psychiatry and Mental Health*, 3, Article 7. https://doi.org/10.1186/1753-2000-3-7

Elder, G. H., Eccles, J. S., Ardelt, M., & Lord, S. (1995). Inner-city parents under economic pressure: Perspectives on the strategies of parenting. *Journal of Marriage and Family*, 57(3), 771–784. https://doi.org/10.2307/353931.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5.

Erskine, H. E., Baxter, A. J., Patton, G., Moffitt, T. E., Patel, V., Whiteford, H. A., & Scott, J. G. (2017). The global coverage of prevalence data for mental disorders in children and adolescents. *Epidemiology and Psychiatric Sciences*, 26(4), 395–402. https://doi.org/10.1017/S2045796015001158.

Fabiano, G. A. (2007). Father participation in behavioral parent training for ADHD: Review and recommendations for increasing inclusion and engagement. *Journal of Family Psychology*, 21(4), 683–693. https://doi.org/10.1037/0893-3200.21.4.683.

Freeman, K. A., & DeCourcey, W. (2007). Further analysis of the discriminate validity of the Parenting Scale. *Journal of Psychopathology and Behavioral Assessment*, 29(3), 169–176. https://doi.org/10.1007/s10862-006-9040-y.

Gardner, F., Hutchings, J., Bywater, T., & Whitaker, C. (2010). Who benefits and how does it work? Moderators and mediators of outcome in an effectiveness trial of a parenting intervention. *Journal of Clinical Child and Adolescent Psychology*, 39(4), 568–580. https://doi.org/10.1080/15374416.2010.486315.

GBD 2019 Risk Factors Collaborators. (2020). Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, 396(10258), 1223–1249. https://doi.org/10.1016/S0140-6736(20)30752-2.

Griner, D., & Smith, T. B. (2006). Culturally adapted mental health intervention: A meta-analytic review. *Psychotherapy*, 43(4), 531–548. https://doi.org/10.1037/0033-3204.43.4.531.

Henninger, W. R., IV, & Luze, G. (2014). Poverty, caregiver depression and stress as predictors of children's externalizing behaviours in a low-income sample. *Child & Family Social Work*, 19(4), 467–479. https://doi.org/10.1111/cfs.12046.

Hillis, S., Mercy, J., Amobi, A., & Kress, H. (2016). Global prevalence of past-year violence against children: A systematic review and minimum estimates. *Pediatrics*, 137(3), e20154079. https://doi.org/10.1542/peds.2015-4079.

Hoffman, K. T., Marvin, R. S., Cooper, G., & Powell, B. (2006). Changing toddlers' and preschoolers' attachment classifications: The Circle of Security intervention. *Journal of Consulting and Clinical Psychology*, 74(6), 1017–1026. https://doi.org/10.1037/0022-006X.74.6.1017.

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

Hurley, K. D., Huscroft-D'Angelo, J., Trout, A., Griffith, A., & Epstein, M. (2014). Assessing parenting skills and attitudes: A review of the psychometrics of parenting measures. *Journal of Child and Family Studies*, 23(5), 812–823. https://doi.org/10.1007/s10826-013-9733-2.

Irvine, A. B., Biglan, A., Smolkowski, K., & Ary, D. V. (1999). The value of the Parenting Scale for measuring the discipline practices of parents of middle school children. *Behaviour Research and Therapy*, 37(2), 127–142. https://doi.org/10.1016/S0005-7967(98)00114-4.

Jansen, E., Frantz, I., Hutchings, J., Lachman, J., Williams, M., Taut, D., Baban, A., Raleva, M., Lesco, G., Ward, C., Gardner, F., Fang, X., Heinrichs, N., & Foran, H. M. (2021). Preventing child mental health problems in southeastern Europe: Feasibility study (phase 1 of MOST framework). *Family Process*. Advance online publication. https://doi.org/10.1111/famp.12720

Juffer, F., Struis, E., Werner, C., & Bakermans-Kranenburg, M. J. (2017). Effective preventive interventions to support parents of young children: Illustrations from the Video-feedback Intervention to promote Positive Parenting and Sensitive Discipline (VIPP-SD). *Journal of Prevention & Intervention in the Community*, 45(3), 202–214. https://doi.org/10.1080/10852352.2016.1198128.

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology*, 36(4), 567–589. https://doi.org/10.1007/s10802-007-9201-9.

Karazsia, B. T., Van Dulmen, M. H. M., & Wildman, B. G. (2008). Confirmatory factor analysis of Arnold et al.'s parenting scale across race, age, and sex. *Journal of Child and Family Studies*, 17(4), 500–516. https://doi.org/10.1007/s10826-007-9170-1.

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., … Rahman, A. (2011). Child and adolescent mental health worldwide: Evidence for action. *The Lancet*, *378*(9801), 1515–1525. https://doi.org/10.1016/S0140-6736(11)60827-1.

Kliem, S., Lohmann, A., Mößle, T., Foran, H. M., Hahlweg, K., Zenger, M., & Brähler, E. (2019). Development and validation of a Parenting Scale short form (PS-8) in a representative population sample. *Journal of Child and Family Studies*, *28*(1), 30–41. https://doi.org/10.1007/s10826-018-1257-3.

Lachman, J. M., Heinrichs, N., Jansen, E., Brühl, A., Taut, D., Fang, X., Gardner, F., Hutchings, J., Ward, C. L., Williams, M. E., Raleva, M., Bǎban, A., Lesco, G., & Foran, H. M. (2019). Preventing child mental health problems through parenting interventions in Southeastern Europe (RISE): Protocol for a multi-country cluster randomized factorial study. *Contemporary Clinical Trials*, *86*, Article 105855. https://doi.org/10.1016/j.cct.2019.105855

Lansford, J. E., Zietz, S., Putnick, D. L., Deater-Deckard, K., Bradley, R. H., Costa, M., Esposito, G., & Bornstein, M. H. (2020). Men's and women's views on acceptability of husband-to-wife violence and use of corporal punishment with children in 21 low- and middle-income countries. *Child Abuse & Neglect*, *108*, Article 104692. https://doi.org/10.1016/j.chiabu.2020.104692

Levav, I., Jacobsson, L., Tsiantis, J., Kolaitis, G., & Ponizovsky, A. (2004). Psychiatric services and training for children and adolescents in Europe: Results of a country survey. *European Child & Adolescent Psychiatry*, *13*(6), 395–401. https://doi.org/10.1007/s00787-004-0427-6.

Lorber, M. F., & Slep, A. M. S. (2018). The reliability paradox of the Parent-Child Conflict Tactics Corporal Punishment subscale. *Journal of Family Psychology*, *32*(1), 145–150. https://doi.org/10.1037/fam0000307.

Lorber, M. F., Xu, S., Slep, A. M. S., Bulling, L., & O'Leary, S. G. (2014). A new look at the psychometrics of the Parenting Scale through the lens of item response theory. *Journal of Clinical Child and Adolescent Psychology*, *43*(4), 613–626. https://doi.org/10.1080/15374416.2014.900717.

MacKenzie, M. J., Nicklas, E., Brooks-Gunn, J., & Waldfogel, J. (2015). Spanking and children's externalizing behavior across the first decade of life: Evidence for transactional processes. *Journal of Youth and Adolescence*, *44*(3), 658–669. https://doi.org/10.1007/s10964-014-0114-y.

McEachern, A. D., Dishion, T. J., Weaver, C. M., Shaw, D. S., Wilson, M. N., & Gardner, F. (2012). Parenting Young Children (PARYC): Validation of a self-report parenting measure. *Journal of Child and Family Studies*, *21*(3), 498–511. https://doi.org/10.1007/s10826-011-9503-y.

McWayne, C. M., Mattis, J. S., Green Wright, L. E., Limlingan, M. C., & Harris, E. (2017). An emic, mixed-methods approach to defining and measuring positive parenting among low-income Black families. *Early Education and Development*, *28*(2), 182–206. https://doi.org/10.1080/10409289.2016.1208601.

Muthén, L. K., & Muthén, B. O. (2019). *Mplus user's guide* (8th ed.).

Naumann, S., Bertram, H., Kuschel, A., Heinrichs, N., Hahlweg, K., & Döpfner, M. (2010). Der Erziehungsfragebogen (EFB) ein Fragebogen zur erfassung elterlicher verhaltenstendenzen in schwierigen Erziehungssituationen [The Erziehungsfragebogen (EFB) a questionnaire for assessing parental behavioural tendencies in difficult parenting situation]. *Diagnostica*, *56*(3), 144–157. https://doi.org/10.1026/0012-1924/a000018.

Neece, C. L., Green, S. A., & Baker, B. L. (2012). Parenting stress and child behavior problems: A transactional relationship across time. *American Journal on Intellectual and Developmental Disabilities*, *117*(1), 48–66. https://doi.org/10.1352/1944-7558-117.1.48.

Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical Child and Family Psychology Review*, *11*(3), 114–144. https://doi.org/10.1007/s10567-008-0033-0.

Phares, V., Lopez, E., Fields, S., Kamboukos, D., & Duhig, A. M. (2005). Are fathers involved in pediatric psychology research and treatment? *Journal of Pediatric Psychology*, *30*(8), 631–643. https://doi.org/10.1093/jpepsy/jsi050.

Pinquart, M. (2017). Associations of parenting dimensions and styles with internalizing symptoms in children and adolescents: A meta-analysis. *Marriage & Family Review*, *53*(7), 613–640. https://doi.org/10.1080/01494929.2016.1247761.

Prinzie, P., Onghena, P., & Hellinckx, W. (2007). Reexamining the Parenting Scale: Reliability, factor structure, and concurrent validity of a scale for assessing the discipline practices of mothers and fathers of elementary-school-aged children. *European Journal of Psychological Assessment*, *23*(1), 24–31. https://doi.org/10.1027/1015-5759.23.1.24.

Pritchett, R., Kemp, J., Wilson, P., Minnis, H., Bryce, G., & Gillberg, C. (2011). Quick, simple measures of family relationships for use in clinical practice and research. A systematic review. *Family Practice*, *28*(2), 172–187. https://doi.org/10.1093/fampra/cmq080.

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Social Science & Medicine*, *90*, 24–31. https://doi.org/10.1016/j.socscimed.2013.04.026.

Reitman, D., Currier, R. O., Hupp, S. D. A., Rhode, P. C., Murphy, M. A., & O'Callaghan, P. M. (2001). Psychometric characteristics of the Parenting Scale in a head start population. *Journal of Clinical Child & Adolescent Psychology*, *30*(4), 514–524. https://doi.org/10.1207/S15374424JCCP3004_08.

Rhoades, K. A., & O'Leary, S. G. (2007). Factor structure and validity of the Parenting Scale. *Journal of Clinical Child and Adolescent Psychology*, *36*(2), 137–146. https://doi.org/10.1080/15374410701274157.

Salari, R., Terreros, C., & Sarkadi, A. (2012). Parenting Scale: Which version should we use? *Journal of Psychopathology and Behavioral Assessment*, *34*(2), 268–281. https://doi.org/10.1007/s10862-012-9281-x.

Sanders, M. R. (2008). Triple P-Positive Parenting Program as a public health approach to strengthening parenting. *Journal of Family Psychology*, *22*(4), 506–517. https://doi.org/10.1037/0893-3200.22.3.506.

Scaramella, L. V., Neppl, T. K., Ontai, L. L., & Conger, R. D. (2008). Consequences of socioeconomic disadvantage across three generations: Parenting behavior and child externalizing problems. *Journal of Family Psychology*, *22*(5), 725–733. https://doi.org/10.1037/a0013190.

Steele, R. G., Nesbitt-Daly, J. S., Daniel, R. C., & Forehand, R. (2005). Factor structure of the Parenting Scale in a low-income african american sample. *Journal of Child and Family Studies*, *14*(4), 535–549. https://doi.org/10.1007/s10826-005-7187-x.

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., … Dees, J. E. M. E. G. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior*, *14*(1), 13–29. https://doi.org/10.1016/j.avb.2006.03.006.

Stormshak, E. A., Bierman, K. L., McMahon, R. J., & Lengua, L. J. (2000). Parenting practices and child disruptive behavior problems in early elementary school. *Journal of Clinical Child Psychology*, *29*(1), 17–29. https://doi.org/10.1207/S15374424jccp2901_3.

Sundell, K., Beelmann, A., Hasson, H., & von Thiele Schwarz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child and Adolescent Psychology*, *45*(6), 784–796. https://doi.org/10.1080/15374416.2015.1020540.

The World Bank. (n.d.). Data for Romania, Moldova, North Macedonia. Retrieved December 27, 2021, from https://data.worldbank.org/

Thomas, R., & Zimmer-Gembeck, M. J. (2007). Behavioral outcomes of Parent-Child Interaction Therapy and Triple P-Positive Parenting Program: A review and meta-analysis. *Journal of Abnormal Child Psychology*, *35*(3), 475–495. https://doi.org/10.1007/s10802-007-9104-9.

UNICEF. (2014). *A statistical snapshot of violence against children in Central and Eastern Europe and the Commonwealth of Independent States (CEE/CIS)*. https://data.unicef.org/resources/statistical-snapshot-violence-children-central-eastern-europe-commonwealth-independent-states-ceecis/