AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Family Psychology

ISSN: 0893-3200

2025, Vol. 39, No. 1, 107–120
https://doi.org/10.1037/fam0001271

Revision of the Multidimensional Assessment of Parenting Scale (MAPS) in an Ethnoracially Diverse National Sample

Violeta J. Rodriguez[1], Anne E. Shaffer[2], and Justin Parent[3]
[1] Department of Psychology, University of Illinois Urbana–Champaign
[2] Department of Psychology, University of Georgia
[3] Department of Psychology, University of Rhode Island

Parenting predicts a wide range of developmental outcomes. Thus, improved assessment of parenting has critical implications for screening into services, identifying additional needs during interventions, and monitoring intervention progress. While there is rising interest in family research focusing on racial and ethnic diversity, there is an unmet need for valid, reliable, and equitable assessments of parenting in racially and ethnically diverse populations. The Multidimensional Assessment of Parenting Scale (MAPS) is a measure of parenting with strong psychometric properties. Despite its high utility, the MAPS is limited in that it was developed in a predominantly White sample (87%). Yet, minoritized racial and ethnic groups currently comprise 40% of the U.S. population. Therefore, this project extended the utility of the MAPS to racially and ethnically diverse families by (a) refining MAPS items using $n = 100$ cognitive interviews with parents from racially and ethnically diverse backgrounds to enhance racial and ethnic representation, (b) quantitatively establishing the Revised MAPS factor structure, and (c) assessing reliability, validity, and measurement invariance by race and ethnicity among $N = 1,699$ parents (25.2% non-Hispanic White, 22.2% non-Hispanic Asian, 23.8% non-Hispanic Black, and 25.6% Hispanic). Following revisions guided by cognitive interviews to improve and add items, the factor structure of the MAPS was replicated in the Revised MAPS. Further, measurement invariance, reliability, and validity of the Revised MAPS were partially supported. Using a multimethod approach, this revision increases the utility of the MAPS by extension to this historically underrepresented segment of the U.S. population.

Keywords: mixed methods, parenting, assessment, psychometrics, ethnoracial minorities

Supplemental materials: https://doi.org/10.1037/fam0001271.supp

Parenting significantly influences children's development, with both negative and positive parenting robustly predicting parent and child well-being, including parent and child psychopathology (Rose et al., 2018). Thus, the development of psychometrically robust assessments for evaluating parenting is essential. However, assessing parenting is complex due to its multifaceted and dynamic nature across contexts and developmental stages, and influenced by sociodemographic and cultural factors (Bornstein, 2013; Lindhiem & Shaffer, 2017; Rodriguez, La Barrie, et al., 2023). The field also faces challenges in psychometric validation, with many parenting measures lacking robust psychometric data, highlighting a need for high-quality parenting assessments (Hurley et al., 2014). Based on these findings, Hurley et al. (2014) classified the state of psychometric evidence of parenting measures as "dismal" (p. 820). Thus, the development of valid and reliable assessments is critical to the advancement of parenting research.

The Multidimensional Assessment of Parenting Scale (MAPS; Parent & Forehand, 2017) partially addresses these challenges by

This article was published Online First September 30, 2024.
Chrystyna D. Kouros served as action editor.
Violeta J. Rodriguez https://orcid.org/0000-0001-8543-2061
Anne E. Shaffer https://orcid.org/0000-0002-2365-8724
Justin Parent https://orcid.org/0000-0002-1659-8188
Analysis code, data, and other supplemental material can be accessed on the Open Science Framework at https://osf.io/a84u2. Violeta J. Rodriguez was affiliated with the University of Georgia during the data collection and analysis phases but has since transitioned to the University of Illinois Urbana–Champaign. This research was part of Violeta J. Rodriguez's dissertation project, with findings shared with a select group of researchers at the University of Georgia as part of her dissertation defense. This study and Violeta J. Rodriguez were supported by a grant from the National Institute of Mental Health under Award Number R36MH127838, and by DP5 Early Independence Award from the Office of the Director under Award Number DP5-OD036508, both from the National Institutes of Health. Thus, as part of the funding process, the study aims were publicly posted on https://reporter.nih.gov prior to data collection. Violeta J. Rodriguez was also funded by a Ford Foundation Predoctoral Fellowship (administered by the National Academies of Sciences, Engineering, and Medicine), by a P.E.O. Scholar Award. The funders had no involvement in the study design, data collection, analysis, decision to publish, or article preparation.

Violeta J. Rodriguez played a lead role in conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, and writing–original draft. Anne E. Shaffer played a lead role in supervision and writing–review and editing. Justin Parent played a lead role in conceptualization and a supporting role in investigation, supervision, and writing–review and editing.

Correspondence concerning this article should be addressed to Violeta J. Rodriguez, Department of Psychology, University of Illinois Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820, United States. Email: vjrodrig@illinois.edu

integrating items from several validated measures, covering a broad developmental range (3–17), and showing psychometric strengths. Importantly, the measure's factor structure aligns with Schaefer's (1959) model, including proactive parenting, positive reinforcement, warmth, supportiveness, hostility, physical control, and lax control. In compiling items from various measures to compose the MAPS, a priority was placed on reducing low-performing, redundant, and repetitive items, while retaining items and subscales commonly targeted in prevention and intervention efforts and those consistent with Schaefer's (1959) circumplex model of maternal behavior. Therefore, the MAPS has been responsive to calls over the past 2 decades to construct valid and reliable measures of parenting that capture a wide developmental span (Hurley et al., 2014; Lindhiem & Shaffer, 2017; Locke & Prinz, 2002). While interview and observational measures of parenting exist, self-report measures are the most common and most widely studied, due to being accessible and feasible, allowing for the collection of valuable subjective insights from parents without requiring extensive resources (Locke & Prinz, 2002). Yet, the representation of minoritized racial and ethnic groups across all types of measures remains limited, affecting the generalizability and equity of measures like the MAPS, particularly given the shifting demographics of the U.S. population (Locke & Prinz, 2002; Rodriguez, La Barrie, et al., 2023; United States Census Bureau, 2019). A lack of measurement invariance (MI) has been identified in the MAPS' negative parenting dimensions, though this prior study is limited by a predominantly non-Hispanic White sample (66.9%). Psychometric properties such as MI are critical for ensuring that parenting scales are valid across diverse groups, yet this has been insufficiently addressed, limiting the applicability of existing scales (Rodriguez, La Barrie, et al., 2023).

Another important consideration emerging from a recent systematic review on MI of parenting measures across racial and ethnic groups is that no studies used qualitative approaches (Rodriguez, La Barrie, et al., 2023). Eliciting feedback from parents in developing measures is particularly important when working with communities historically excluded from research. One method of doing this may be through cognitive interviews, which involves the process of learning how participants interpret scale items through interviews to improve their quality (Willis, 2004). Cognitive interviews can also guide revisions by eliciting feedback from participants regarding challenges when completing the scale.

The Present Study

Historically, most research on parenting has been conducted among White, middle-class families, providing a limited conceptualization of parenting and raising questions about the universal applicability of parenting models. Although more attention has recently been given to racial and ethnic diversity in family research, this has not extended to psychometric research (Rodriguez, La Barrie, et al., 2023; Rodriguez, Cadet, et al., 2023). As such, psychometric properties of parenting measures with minoritized racial and ethnic groups have received limited consideration, contributing to the absence of measures of parenting for use in racially and ethnically diverse populations. Further, preliminary evidence shows that in a combined clinical and community sample of parents, though positive parenting domains can be equivalently assessed by the MAPS across racial and ethnic groups, negative parenting may not (Rodriguez, Cadet, et al., 2023). However, this prior study is limited by a mostly non-Hispanic White sample (66.9%). Thus, more diverse samples and multimethod approaches are needed to assess and minimize measurement bias arising from a measure that was developed in a predominantly White sample and based on research and theories emerging from Eurocentric perspectives. Therefore, the present study combines qualitative and quantitative methods to revise the MAPS for diverse populations in a national racially and ethnically diverse sample powered to test MI, with three goals: (1) refine the MAPS dimensions, items, and item formats using qualitative interviews; (2) establish the factor structure of the Revised MAPS in a racially and ethnically diverse national sample; and (3) establish the reliability, validity, and MI of the Revised MAPS in a diverse sample.

Method

Phase 1: Qualitative Interviews

Ethical Approval

Ethical approval from the University of Georgia Institutional Review Board was obtained.

Participant Eligibility

Eligible parents for the interviews and survey were at least 18 years old, parenting at least one child aged 3–17, and living with their target child at least 50% of the time, consistent with the MAPS development and validation sample. If the parent had more than one child, they were instructed to choose the one with the most recent birthday.

Qualitative Interviews

Cognitive interviews ($n = 100$) were conducted to obtain feedback on the MAPS in English. Cognitive interviews capture participants' thought processes when engaging with survey items, probing how participants understand and interpret items (Willis, 2004).

Recruitment. Participants were recruited by race and ethnicity ($n = 25$ non-Hispanic White, $n = 25$ non-Hispanic Black, $n = 25$ non-Hispanic Asian, and $n = 25$ Hispanic regardless of race). To recruit participants, paid, targeted social media advertisements on Facebook and Instagram across the U.S. were used. Flyers included interview duration (1 hr), compensation (USD$50), an email address to contact research personnel, and a link to an eligibility survey where parents could send their information to research personnel to be screened.

Procedure. Eligible participants received surveys distributed through Qualtrics, which displayed the informed consent letter. After completion of the survey and agreement to participate, participants were sent a link to schedule their interview with the research team using Calendly, which automatically emailed a Zoom password-protected meeting identifier.

Interviews and Transcription. Interviews were conducted online via Zoom video conferencing. Researchers shared their screens to present a slide deck with each item and responses displayed on a slide. Researchers then read each item aloud and asked the parent to respond; their response was recorded in Qualtrics by the researcher. Following parents' responses, parents were prompted to elaborate on their response, interpretation of the item, and any points of confusion using cognitive interviewing techniques (Willis, 2004). Questions in cognitive interviews used in previous measurement research guided questions and stem (Willis, 2004). Participants were then asked open-ended questions to obtain suggestions of topics that may be additionally relevant to the scale, such as, "Describe some strategies you use to raise your child(ren)." Most interviews (n = 67) were conducted by the first author and others by senior undergraduate research assistants (n = 33). All interviews were audio-recorded, and https://Otter.ai was used to transcribe recordings. Researcher personnel reviewed transcripts to ensure accurate transcription.

Analysis

Descriptive Statistics. To characterize the sample, univariate analyses were used, including means and standard deviations for continuous variables and frequencies for categorical variables.

Qualitative Analyses. Qualitative content analysis was used to identify patterns in qualitative data derived from transcripts. NVivo was used for coding, sorting, and quantification of the data about the item response process by item. The Patient-Reported Outcome Measurement Information System Qualitative Item Review was then used to analyze and achieve consensus on item removal or revision, focusing on item phrasing, response options, and literacy demands, revising items with ambiguity or cognitive difficulty to large proportions of participants (DeWalt et al., 2007). NVivo was used to organize thematic issues for each of the items to guide item revision. *A priori*, it was decided that major revisions (i.e., more than one or two words) would be made to items if more than 30% of participants across all groups reported ambiguity, cognitive difficulty, or confusion related to an item, which was consistent with rates used in previous studies using cognitive interviews (e.g., Matza et al., 2015). *A posteriori*, as data was collected but before final decisions were made, it was decided that one- or two-word or clarifying revisions (e.g., providing examples) would be made if it was reported by more than 30% of one specific group (e.g., 30% of *all* fathers), or 15% by the overall group.

Phase 2: Quantitative Survey

Ethical Approval

Similar to Phase 1, ethical approval for Phase 2 was obtained from the (University of Georgia) Institutional Review Board.

Participant Eligibility

The eligibility criteria for Phase 2 were identical to Phase 1. Participants participating in the survey were a separate group of participants from those participating in the cognitive interviews.

National Survey

The Revised MAPS and other measures were distributed as surveys by QualtricsXM, a data collection and management company. Participants in Qualtrics panels are nationally representative samples of adults who volunteer to participate in surveys online in exchange for compensation. Qualtrics contacted their prescreened participant panels to assess eligibility and emailed the survey to participants to complete via a link, then disbursed compensation. Recruitment ran from November to December 2022. In total, N = 1,699 eligible respondents completed the survey in English. For quality purposes, only eligible participants who correctly and appropriately answered attention and quality checks embedded throughout the survey were included in the final data set. Specifically, the survey was piloted in n = 150 participants initially and an average was taken of their survey completion time before data collection was resumed; these pilot participants were included in the final sample. Participants whose survey completion time was faster than 1.5 *SD*s from the pilot average were excluded (n = 35).

Measures

Sociodemographic Characteristics. Participants were asked to complete a questionnaire about their age, gender, race, ethnicity, educational attainment, relationship status, employment status, source of income, personal and household income, living arrangements (i.e., time spent with child), the number of children, and children's gender and age.

Multidimensional Assessment of Parenting Scale. The 34-item MAPS (Parent & Forehand, 2017) is composed of seven subscales of parenting within two broadband domains (i.e., positive and negative parenting). The original MAPS has a 16-item broadband positive parenting subscale composed of items related to proactive parenting, positive reinforcement, warmth, and supportive parenting. The original MAPS also includes an 18-item negative parenting subscale addressing hostility, lax control, and physical control. Items are rated on a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*), with greater scores reflecting higher positive or negative parenting, depending on the specified domain or subscale.

Parental Discipline. The nine-item Alabama Parenting Questionnaire (APQ-9; Frick, 1991) was used to measure parental discipline for convergent validity. Subscale scores for positive parenting (ω = .74), inconsistent discipline (ω = .78), and poor supervision (ω = .85) were used.

Parent Symptoms of Depression. The Patient Health Questionnaire–9 (PHQ-9) was used to measure symptoms of depression in parents over the past 2 weeks (Kroenke et al., 2009). Total scores range from 0 to 27 (ω = .91), with severity cutoffs being as follows: no depressive symptoms (0–4), mild depressive symptoms (5–9), moderate depressive symptoms (10–14), moderately severe depressive symptoms (15–19), and severe depressive symptoms (20–27).

Parent Symptoms of Anxiety. The General Anxiety Disorder–7 (GAD-7) was used to measure symptoms of generalized anxiety disorder in the past 2 weeks (Spitzer et al., 2006). Total scores range from 0 to 21; continuous scores were used (ω = .92). To describe severity levels, we used the following cutoffs: minimal anxiety (0–4), mild anxiety (5–9), and moderate anxiety (10–14).

Child Symptoms of Psychopathology. Parents completed the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), a 25-item behavioral screening questionnaire that assesses conduct problems, hyperactivity-inattention, emotional symptoms, peer problems, and prosocial behavior. For parents of target children ages 2–4, reliability was acceptable (ω = .79). For parents of target children ages 5–11 and 12–17, reliability was also acceptable (ω = .83). To describe the percentage of children with abnormal levels of psychopathology (per parent report), a cutoff of 17 was used.

Analysis

Factor Structure

Following revisions of the MAPS and data collection using national surveys, the second stage of analysis was to establish the factor structure of the Revised MAPS in a national sample oversampling minoritized racial and ethnic groups. Given that only one item was added to the original 34-item MAPS, the original factor structure was estimated using confirmatory factor analysis (CFA). Five different competing CFA models were estimated consistent with previous research (Parent & Forehand, 2017). Graphical representations of these models are provided in Supplemental Figure 1, with greater detail on these models provided in the Supplemental Material. Given that CFA models can be overly simplistic, limited, and idealistic in their assumption of "pure factors" that do not correlate and in which cross-loadings are constrained to zero, five competing exploratory structural equation models (ESEMs) were tested for which graphical representations are provided in Supplemental Figure 2, as in the original MAPS development and validation study (Parent & Forehand, 2017), with greater detail on these ESEMs also presented in the Supplemental Material. Across all models, maximum likelihood estimation with robust standard errors was used to account for nonnormality. ESEMs used an oblique rotation with target loading values close to zero. Model fit across these models was evaluated using χ^2 tests, comparative fit index (CFI), Tucker–Lewis index, root-mean-square error of approximation (RMSEA), and standardized root-mean-square residual. A CFI value $\geq$.95 indicates a "good" fit of the model to the observed data (Brown, 2014; Kline, 2015; van Zyl & Ten Klooster, 2022). Similarly, an RMSEA value $\leq$.05 indicates a close fit of the model to the data (Brown, 2014; Kline, 2015; van Zyl & Ten Klooster, 2022). Given that these model fit criteria are based on CFA models and not ESEMs, indices of measurement quality were used to further discriminate between models. Specifically, measurement quality was evaluated as standardized factor loadings (λ) >0.35, item uniqueness (>0.10 but <0.90), and standard errors <.04.

MI/Equivalence

The next stage included testing for MI utilizing item response theory (IRT) to estimate item parameters and differential item functioning (DIF). Uniform DIF refers to consistent item performance across all score groups, whereas nonuniform DIF refers to a bias in favor of one group that is not constant across total scores. Items with nonuniform DIF may be indicative of bias and were therefore removed. DIF was tested using an iterative hybrid ordinal logistic regression/IRT approach via likelihood ratio tests using the *lordif* package (Choi et al., 2011). Because likelihood ratio tests can overidentify DIF due to the number of model comparisons (Type 1 error), Monte Carlo simulations were used to determine the empirical thresholds to detect DIF. Monte Carlo simulations generate a distribution of R^2 values free of DIF; the highest thresholds are determined from this distribution. R^2 values greater than these thresholds were classified as DIF, as a second method to confirm results. Pseudo-R^2 differences were used as effect sizes to describe DIF, and graphical representations of DIF impact provided by *lordif* were used to determine the need for further removal of DIF items.

Reliability and Validity

The next stage was to establish the reliability, validity, and MI of the Revised MAPS in the national sample. To examine the reliability, McDonald's ω was used. Correlations between Revised MAPS factors, related measures, and similar measures (APQ) were used to establish predictive and convergent validity. Though the APQ lacked MI in previous research (Shaffer et al., 2022), no other relevant multidimensional parenting scales, to our knowledge, had been tested for MI by ethnoracial identity (Rodriguez, La Barrie, et al., 2023). Thus, we explored correlations between the Revised MAPS and the APQ to test the scale's convergent validity. To test for predictive validity, we examined associations with parent and child psychopathology, as a wealth of literature implicates parenting in parent and child psychopathology and vice versa (e.g., Rose et al., 2018).

Differential Predictive Validity

Multiple regression analyses with robust standard errors were conducted to assess the differential predictive validity of each MAPS subscale for child psychopathology (SDQ), parent symptoms of depression (PHQ-9), and anxiety (GAD-7) using R packages *lmtest* and *sandwich*. Interactions (e.g., MAPS Subscale × Ethnoracial Identity) were included in the models to test for differential predictive validity across ethnoracial identity.

Results

Phase 1: Qualitative Interviews

Sociodemographic Characteristics of Parents in Interviews

A total of $N = 100$ interviews were completed with parents. On average, parents were 37.35 years old ($SD = 6.90$); 51% of parents were mothers, and 49% were fathers (see Table 1).

Qualitative Interviews and Related Revisions

The cognitive interviewing process provided rich qualitative data about items regarding language, comprehensibility, ambiguity, and relevance of each item, which are reported in detail in the Supplemental Material. A summary of salient topics of parents' responses is provided in Supplemental Table 1, with specific original and revised items reported in Supplemental Table 2.

Overall, to address the 36% of respondents who sought a broader definition of affection (Item 1), adjustments were made to encompass a wider range of affectionate actions, acknowledging cultural distinctions in the expression of affection. Items with potential cultural implications, such as those related to discipline strategies (Item 5, where 56% expressed confusion, and Items 14 and 25, with nearly half questioning the term "physical punishment"), were refined to include explicit examples or to simplify language to capture a broader spectrum of parenting. Terms that risked misinterpretation due to cultural or linguistic differences were also clarified or substituted, ensuring the scale's relevance and applicability across diverse family structures and cultural backgrounds. Items 6, 15, 19, 22, and 28 were not revised. Item 20 only underwent minimal revisions to accommodate gender neutral pronouns. Items 2, 14, 17, 24, 29, and 32 underwent minimal (i.e., one- or two-word) clarifying revisions. Items 1, 11, 12, 18, 21, 25, and 33 were revised to provide

Table 1
Sociodemographic Characteristics of Parents and Their Children

Variable	Phase 1 (interview) *N* (%) *M* (*SD*)	Phase 2 (survey) *N* (%) *M* (*SD*)
	Parents	
Age	37.35 (6.90)	37.91 (9.20)
Sex		
Male	49 (49%)	876 (51.6%)
Female	51 (51%)	823 (48.4%)
Ethnicity		
Non-Hispanic	75 (75%)	1,264 (74.4%)
Hispanic	25 (25%)	435 (25.6%)
Race		
Asian	26 (26%)	376 (22.1%)
Black	27 (27%)	470 (27.7%)
White	37 (37%)	669 (39.4%)
Other	1 (1%)[a]	172 (10.1%)
Native Hawaiian/Pacific Islander		24 (1.4%)
Native American or Alaskan Native		57 (3.4%)
Prefer not to answer	9 (9%)[a]	
Relationship status		
Never married	15 (15%)	292 (17.2%)
Separated or divorced	9 (9%)	142 (8.4%)
Married	74 (74%)	1,001 (58.9%)
Widowed		14 (0.8%)
In a committed relationship		238 (14.0%)
Prefer not to answer	2 (2%)	12 (0.7%)
Coparenting status		
No	8 (8%)	272 (16.0%)
Yes	91 (91%)	1,395 (82.1%)
Prefer not to answer	1 (1%)	32 (1.9%)
Education		
Postgraduate degree	44 (44%)	260 (15.3%)
Bachelor's degree	36 (36%)	464 (27.3%)
Associate degree	2 (2%)	210 (12.4%)
Partial college	11 (11%)	309 (18.2%)
High school or GED	6 (6%)	413 (24.3%)
Partial high school	1 (1%)	43 (2.5%)
Employment		
Not employed outside the home	17 (17%)	343 (20.2%)
Work part-time	20 (20%)	253 (14.9%)
Work full-time	62 (62%)	1,076 (63.3%)
Prefer not to answer	1 (1%)	27 (1.6%)
Household yearly income		
Less than USD$10,000	3 (3%)	137 (8.1%)
USD$10,000–USD$30,000	9 (9%)	218 (12.8%)
USD$30,000–USD$50,000	13 (13%)	272 (16.0%)
USD$50,000–USD$70,000	11 (11%)	273 (16.1%)
USD$70,000–USD$90,000	14 (14%)	218 (12.8%)
USD$90,000–USD$100,000	5 (5%)	166 (9.8%)
Over USD$100,000/year	40 (40%)	374 (22.0%)
Prefer not to answer	5 (5%)	41 (2.4%)
	Target child	
Age	7.01 (3.92)	9.75 (4.37)
Sex		
Male	56 (56%)	886 (52.1%)
Female	43 (43%)	808 (47.6%)
Intersex	1 (1%)	5 (0.3%)

Note. GED = General Educational Development; USD = United States Dollar.
[a] All participants who identified as "Other" race or who responded as "Prefer not to answer" to race identified as Hispanic.

clearer examples of behavior or to address points of confusion and ambiguity. Items 3 (removed "for misbehavior"), 4 (added "when I'm irritated with them"), 5 (removed "with little or no justification"), 8 (removed "or shout"), 9 (removed "talks me out of"), 10 (removed "show respect"), 13 (replaced "explode in anger" with "get very angry at"), 16 (replaced "lose my temper" with "lose control of my anger"), and 31 (removed "because other things I have tried have not worked") had more than two words removed or added to provide clarifications. Items 23, 26, 27, 30, and 34 underwent substantial revisions for simplification. A final MAPS with response options is presented in Supplemental Material 2.

One item was added to include verbal forms of affection given that 48% of parents mentioned verbal warmth or affection in response to warmth items. With regard to other suggestions from parents, qualitative insights highlighted other areas that they would have suggested for assessment. These included clear and logical communication, recognizing and praising children's efforts, adapting disciplinary methods to children's emotional and developmental needs, and consistency in parenting. Physical affection, setting safety boundaries, respecting and encouraging children's opinions, and managing tantrums, especially in public, were also discussed. However, these were deemed to be addressed by the assessment or to be outside of the scope of a self-report scale (e.g., parenting in various contexts).

Phase 2: Quantitative Survey

Sociodemographic Characteristics of Parents in National Sample

A total of $N = 1{,}699$ parents were recruited in the national sample. On average, parents were 37.91 ($SD = 9.20$) years of age; 48.4% of parents were mothers (see Table 1), with comparisons by ethnoracial identity in Table 2. Importantly, all 50 states were represented in the survey, with the greatest proportion of parents being from California ($n = 219$; 12.9%), Texas ($n = 159$; 9.4%), Florida ($n = 122$; 7.2%), New York ($n = 110$; 6.5%), and Georgia ($n = 84$; 4.9%). No missing data was present in the parenting and psychopathology measures, likely due to reminder prompts when items were missed. The instances in which "Prefer not to answer" was selected for demographics are reported in Table 1.

Severity of Parent and Child Mental Health Symptoms

The distribution of participants across the PHQ-9 severity categories was as follows: no depressive symptoms (0–4; $n = 771$, 46.9%), mild depressive symptoms (5–9; $n = 387$, 23.5%), moderate depressive symptoms (10–14; $n = 229$, 13.9%), moderately severe depressive symptoms (15–19; $n = 176$, 10.7%), and severe depressive symptoms (20–27; $n = 81$, 4.9%). The GAD-7 severity categories were distributed as follows: minimal anxiety (0–4; $n = 1{,}330$, 80.9%), mild anxiety (5–9; $n = 205$, 12.5%), and moderate anxiety (10–14; $n = 109$, 6.6%). In terms of the SDQ total scores, 85.8% of the parents reported SDQ symptoms greater than 17.

Exploratory and Confirmatory Factor Analyses of Revised MAPS

To confirm the dimensionality of the scale before tests of MI, confirmatory and exploratory factor analytical models were estimated and summarized in Table 3. For ease of interpretation, these models are also visually depicted in Supplemental Figures 1 and 2. Though ESEMs generally provided a better fit to data, measurement quality indices were generally poor for ESEMs. Specifically, several factor loadings across all ESEMs were consistently <.35 for positive parenting subscales, with substantially lower factor loadings for reinforcement (e.g., $\lambda = .12$ to $\lambda = .27$) and supportive (e.g., $\lambda = .07$ to $\lambda = .20$) parenting practices subscales. Similarly, several standard errors across most ESEMs were >.04. However, Model 1, which was estimated using a CFA model with seven first-order factors, demonstrated acceptable to excellent fit ($\chi^2 = 2086$, $df = 539$, CFI = 0.93, Tucker–Lewis index = 0.92, RMSEA = 0.041, standardized root-mean-square residual = 0.049, Akaike information criterion = 150,408, Bayesian information criterion = 151,094, adjusted Bayesian information criterion = 150,693) with excellent measurement quality indices, ranging from $\lambda = .55$ to $\lambda = .88$, with target items loading significantly on their respective factors. Given that there is substantial theoretical and empirical support to retain the reinforcement and supportive subscales in the MAPS, Model 1 was retained as the best model. As such, other CFA and ESEMs were not considered further.

MI by Race and Ethnicity

After confirming the factor structure of the Revised MAPS, tests of DIF by race and ethnicity were conducted using ordinal logistic regression model likelihood ratio tests via *lordif* (Choi et al., 2011). One single variable comprised of ethnoracial groups was used to test for DIF by race and ethnicity given that (a) a substantial number of Hispanic participants did not report on race; (b) testing for DIF by race and ethnicity separately would double the number of comparisons from 210 (likelihood ratio tests for overall DIF, nonuniform DIF, and uniform DIF per item [35]) comparisons to 420 comparisons; (c) if an overall test of DIF was significant, individual group pairwise comparisons could be conducted; and (d) there is greater Type 1 error control when group sizes are equal when testing for DIF (González-Romá et al., 2006).

No items in the MAPS Proactive Parenting, Reinforcement, Supportiveness, and Warmth subscales emerged as having nonuniform DIF. However, several items emerged as having nonuniform DIF in the MAPS Hostility, Lax Control, and Physical Control subscales, which were all in the original Negative Parenting practices broadband scale (described in Supplemental Material). To control for Type 1 error due to multiple (210) comparisons, results were confirmed via Monte Carlo simulations. Here, *lordif* fits the ordinal logistic regression models/IRT procedures over $n = 1{,}000$ simulated DIF-free data sets and recalculates statistics and magnitude measures. Finally, *lordif* identifies a threshold value that cuts off the most extreme ($\alpha \times 100$)% of each of the statistics and magnitude measures; the identified value for nonuniform DIF was $R^2_{23} = .012$. No pseudo-R^2_{23} was above this threshold. In addition, examining pseudo-R^2 values across these items for uniform and nonuniform DIF, these values were generally negligible ($R^2_{23} = .003$ for Item 32 to $R^2_{23} = .012$ for Item 26). The item with the highest pseudo-R^2 value, Item 26, is plotted in Supplemental Figure 3. Of note, results were identical when controlling for parent sex.

Given that no items were removed from the MAPS, the final model is presented in Supplemental Table 3, and the final, Revised MAPS items are in Supplemental Table 4. The average grade equivalent

Table 2
Sociodemographic Characteristics of Parents and Children by Ethnoracial Group (Phase 2)

Variable	Asian N (%) M (SD)	Black N (%) M (SD)	White N (%) M (SD)	Hispanic N (%) M (SD)
		Parents		
Age	40.31 (9.02)	36.24 (9.26)	39.37 (8.67)	35.91 (9.21)
Sex				
Male	133 (35.3%)	249 (61.6%)	358 (83.6%)	208 (47.8%)
Female	244 (64.7%)	155 (38.4%)	70 (16.4%)	227 (52.2%)
Relationship status				
Never married	23 (6.1%)	143 (35.4%)	38 (8.9%)	74 (17.0%)
Separated or divorced	23 (6.1%)	35 (8.7%)	42 (9.8%)	38 (8.7%)
Married	296 (78.5%)	141 (34.9%)	296 (69.2%)	243 (55.9%)
Widowed	1 (0.3%)	5 (1.2%)	3 (0.7%)	5 (1.1%)
In a relationship	34 (9.0%)	76 (18.8%)	49 (11.4%)	79 (16.1%)
Prefer not to answer	0 (0.0%)	4 (0.0%)	0 (0.0%)	5 (1.1%)
Coparenting status				
No	37 (9.8%)	107 (26.5%)	48 (11.2%)	72 (16.6%)
Yes	329 (87.3%)	293 (72.5%)	375 (87.6%)	356 (81.8%)
Prefer not to answer	11 (2.9%)	4 (1.0%)	5 (1.2%)	7 (1.6%)
Education				
Postgraduate degree	85 (19.9%)	101 (26.8%)	27 (6.7%)	42 (9.7%)
Bachelor’s degree	130 (30.4%)	155 (41.1%)	74 (18.3%)	97 (22.3%)
Associate degree	45 (10.5%)	34 (9.0%)	70 (17.3%)	49 (11.3%)
Partial college	67 (15.7%)	41 (10.9%)	94 (23.3%)	91 (20.9%)
High school or GED	90 (21.0%)	43 (11.4%)	128 (31.7%)	138 (31.7%)
Partial high school	11 (2.6%)	3 (0.8%)	11 (2.7%)	18 (4.1%)
Employment				
Not employed outside the home	59 (13.8%)	88 (23.3%)	96 (23.8%)	86 (19.8%)
Work part-time	37 (8.6%)	67 (17.8%)	71 (17.6%)	68 (15.6%)
Work full-time	327 (76.4%)	213 (56.5%)	232 (57.4%)	274 (63.0%)
Prefer not to answer	5 (1.2%)	9 (2.4%)	5 (1.2%)	7 (1.6%)
Household yearly income	18 (4.2%)	16 (4.2%)	66 (16.3%)	34 (7.8%)
Less than USD$10,000	39 (9.1%)	20 (5.3%)	83 (20.5%)	64 (14.7%)
USD$10,000–USD$30,000	54 (12.6%)	43 (11.4%)	86 (21.3%)	76 (17.5%)
USD$30,000–USD$50,000	70 (16.4%)	51 (13.5%)	70 (17.3%)	75 (17.2%)
USD$50,000–USD$70,000	58 (13.6%)	55 (14.6%)	38 (9.4%)	61 (14.0%)
USD$70,000–USD$90,000	63 (14.7%)	28 (7.4%)	23 (5.7%)	48 (11.0%)
USD$90,000–USD$100,000	125 (29.2%)	144 (38.2%)	32 (7.9%)	68 (15.6%)
Over USD$100,000/year	1 (0.2%)	20 (5.3%)	6 (1.5%)	9 (2.1%)
Prefer not to answer	18 (4.2%)	16 (4.2%)	66 (16.3%)	34 (7.8%)
		Target child		
Age	9.89 (4.40)	9.28 (4.40)	10.23 (4.27)	9.50 (4.32)
Sex				
Male	251 (58.6%)	174 (46.2%)	207 (51.2%)	225 (51.7%)
Female	176 (41.1%)	201 (53.3%)	197 (48.8%)	209 (48.0%)
Intersex	1 (0.2%)	2 (0.5%)	0 (0.0%)	1 (0.2%)

Note. GED = General Educational Development; USD = United States Dollar.

reading level of all items was 6.2, measured using the web-based Readable app (https://app.readable.com), as with the original MAPS (Parent & Forehand, 2017).

Ethnoracial Comparisons

Because MI of the Revised MAPS was supported, ethnoracial comparisons are presented in Table 4. Parent sex, age, income, education, having more than one child, and child age and sex were evaluated as covariates and included if associated with Revised MAPS subscales at $p < .05$. In an analysis of the Revised MAPS subscales across ethnoracial groups, significant differences were observed in several dimensions of parenting. Specifically, analyses of variance revealed significant variances by ethnoracial groups in Hostility, $F(3, 996) = 5.78$, $p < .001$; Physical Control, $F(3, 996) = 11.77$, $p < .01$; and Reinforcement, $F(3, 996) = 4.10$, $p < .01$. Post hoc comparisons with Bonferroni corrections revealed that non-Hispanic White parents reported higher levels of Hostility compared to non-Hispanic Asian, non-Hispanic Black, and Hispanic parents ($p < .001$). Similarly, for Physical Control, non-Hispanic White parents demonstrated higher levels than non-Hispanic Asian, non-Hispanic Black, and Hispanic parents ($p < .01$). Moreover, non-Hispanic Black parents themselves exhibited higher levels of Physical Control in comparison to both non-Hispanic Asian and Hispanic parents, further highlighting significant differences within these groups. Additionally, in terms of Reinforcement, non-Hispanic White

Table 3
Competing Confirmatory and Exploratory Factor Analytical Models

Model	Type	χ^2	*df*	CFI	TLI	RMSEA	SRMR	AIC	BIC	aBIC	Fit	Measurement quality
Confirmatory factor analytical models												
Model 0	Unidimensional first-order factor model	13,137	560	0.42	0.38	0.115	0.183	165,086	165,657	165,323	No	No
Model 1	Seven first-order factor model	2,087	539	0.93	0.92	0.041	0.049	150,408	151,094	150,693	Yes	Yes
Model 2	Second-order factor model	2,570	554	0.91	0.90	0.046	0.119	151,000	151,604	151,251	No	Yes
Model 2b	Third-order factor model	2,844	554	0.89	0.89	0.049	0.138	151,367	151,971	151,618	No	Yes
Model 3	Bifactor model	3,210	525	0.88	0.86	0.055	0.142	151,864	152,626	152,181	No	No
Exploratory factor analytical models												
Model 4	Seven first-order ESEM	865	371	0.98	0.96	0.028	0.015	149,094	150,692	149,758	Yes	No
Model 5	Higher order ESEM	1,805	385	0.95	0.93	0.047	0.072	149,741	151,263	150,374	No	No
Model 6	Bifactor ESEM	950	343	0.98	0.97	0.032	0.013	148,970	150,721	149,698	Yes	No
Model 7	ESEM within CFA	865	371	0.98	0.96	0.028	0.015	149,094	150,692	149,758	Yes	No
Model 8	Hierarchical EwC	996	384	0.97	0.96	0.031	0.027	149,242	150,770	149,877	Yes	No

Note. Model 0 included a unidimensional first-order factor model in which all items were specified to load directly on one factor. Model 1 was a seven first-order factor model in which the seven original MAPS factors were specified and the factors were allowed to freely correlate. Model 2 included a model of parenting as a function of two broadband subscales of positive and negative parenting. Model 2b was a model of parenting as a function of one broad third-order parenting factor with two broadband subscales of positive and negative parenting. Lastly, Model 3 was a bifactor model that was estimated to be comprised of a General Factor representing parenting (where all 35 items were specified to load onto such directly), which is distinct and independent from its seven first-order factors. Model 4 included a seven-factor first-order ESEM, in which the seven MAPS factors were estimated as distinct and related first-order factors. Model 5 was a higher order ESEM (H-ESEM), in which parenting was estimated (via the ESEM-within-CFA framework) as a single, second-order factorial model comprised out of seven first-order factors. Model 6 was a bifactor ESEM, such that parenting is seen as a general factor, that is distinct from its seven independent specific factors of parenting. Lastly, Model 7 was an ESEM-within-CFA model, in which a seven-factor first-order ESEM is reexpressed within a CFA framework using the starting values of the original seven-factor first-order ESEM. Using an ESEM within CFA (EwC) framework, as in the original MAPS validation study, Model 8 was estimated including two broadband subscales of positive and negative parenting including proactive parenting, positive reinforcement, warmth, and supportiveness loading onto broadband positive parenting, and hostility, lax control, and physical control loading onto broadband negative parenting. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; aBIC = adjusted Bayesian information criterion; ESEM = exploratory structural equation modeling; CFA = confirmatory factor analysis; MAPS = Multidimensional Assessment of Parenting Scale.

Table 4
Revised MAPS Subscales by Ethnoracial Groups

	MAPS H	MAPS L	MAPS P	MAPS PP	MAPS R	MAPS S	MAPS W
Ethnoracial identity	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Non-Hispanic White	16.73 (6.42)	16.72 (6.72)	7.91 (4.49)	22.78 (4.36)	16.67 (3.10)	12.65 (2.28)	16.79 (2.91)
Non-Hispanic Asian	16.05 (5.07)	15.95 (5.10)	6.25 (3.30)	22.62 (4.50)	16.97 (2.86)	12.88 (2.01)	16.94 (3.02)
Non-Hispanic Black	14.24 (5.48)	15.03 (5.62)	7.34 (3.55)	23.02 (5.35)	16.62 (3.71)	13.10 (2.46)	17.28 (3.36)
Hispanic	15.08 (5.84)	15.65 (5.75)	6.80 (3.62)	22.45 (5.26)	16.51 (3.75)	12.88 (2.53)	17.07 (3.31)
F test	5.78***[a,b]	0.91	11.77**[a,b,c,d,e]	1.90	4.10**[a,b]	0.53	0.97

Note. MAPS Hostility was adjusted for parent income, parent education, having more than one child, child sex, and child age. MAPS Lax Control was adjusted for parent sex, parent age, parent income, parent education, and child age. MAPS Physical Control was adjusted for parent sex, parent age, parent education, child sex, and child age. MAPS Proactive Parenting, Supportiveness, and Warmth were adjusted for parent sex, parent income, parent education, having more than one child, child sex, and child age. MAPS Reinforcement was adjusted for parent sex, parent income, parent education, having more than one child, and child age. MAPS = Multidimensional Assessment of Parenting Scale; MAPS H = MAPS hostility; MAPS L = lax control; MAPS P = physical control; MAPS PP = proactive parenting; MAPS R = reinforcement; MAPS S = supportiveness; MAPS W = warmth.
[a] Difference between non-Hispanic White and non-Hispanic Black. [b] Difference between non-Hispanic White and Hispanic. [c] Difference between non-Hispanic White and non-Hispanic Asian. [d] Difference between non-Hispanic Asian and non-Hispanic Black. [e] Difference between non-Hispanic Black and Hispanic.
** $p = .01$. *** $p = .001$. All statistically significant differences are Bonferroni-corrected.

parents were found to engage in higher levels compared to non-Hispanic Black and Hispanic parents ($p < .01$). No significant differences were found for Lax Control, Proactive Parenting, Supportiveness, and Warmth (see Table 5).

Reliability and Construct Validity

McDonald's ω values were in the acceptable range for all subscales: $\omega = 0.87$ for Hostility, $\omega = 0.84$ for Lax Control, $\omega = 0.90$ for Physical Control, $\omega = 0.82$ for Proactive Parenting, $\omega = 0.83$ for Reinforcement, $\omega = 0.78$ for Supportiveness, and $\omega = 0.83$ for Warmth. For the full scale, reliability was excellent, $\omega = 0.94$.

After assessing the internal reliability of the Revised MAPS, bivariate correlations were calculated to measure associations with parent psychopathology (PHQ-9, GAD-7) and child psychopathology (SDQ) to assess predictive validity. Additionally, correlations with a general measure of parenting (APQ-9) were used to assess convergent validity. These results were interpreted as evidence of construct validity. Hostility, Lax Control, and Physical Control showed positive correlations with depressive symptoms on the PHQ-9 ($r = 0.457$, $r = 0.383$, $r = 0.431$, respectively) and anxious symptoms on the GAD-7 ($r = 0.352$, $r = 0.267$, $r = 0.339$, respectively). These subscales also correlated positively and moderately with child psychopathology on the SDQ ($r = 0.539$, $r = 0.506$, $r = 0.566$, respectively). In contrast, Proactive Parenting, Reinforcement, Supportiveness, and Warmth showed significant negative correlations with depressive symptoms, although Proactive Parenting did not correlate significantly with the PHQ-9 ($r = -0.019$). These positive parenting dimensions correlated with the APQ Positive Parenting at moderate levels ($r = 0.438$, $r = 0.577$, $r = 0.476$, respectively), providing evidence of convergent validity. Notably, the correlations between negative MAPS subscales (Hostility, Lax Control, Physical Control) and APQ Negative Parenting scales (APQ Inconsistent Discipline and APQ Poor Supervision) were weaker, consistent with expectations for convergent validity patterns. Specifically, Hostility correlated with APQ Inconsistent Discipline ($r = .534$) and APQ Poor Supervision ($r = .490$). Lax Control showed correlations with APQ Inconsistent Discipline ($r = .694$) and APQ Poor Supervision ($r = .547$). Physical Control was associated with APQ Inconsistent Discipline ($r = .490$) and APQ Poor Supervision ($r = .537$).

Differential Predictive Validity

After applying the Bonferroni correction to adjust for multiple comparisons, all corrected *p* values were >.05, and covariates previously specified indicated no statistically significant interaction effects between MAPS subscales and ethnoracial identity on child psychopathology (SDQ total score), parent symptoms of depression (PHQ-9), and anxiety (GAD-7). These findings suggest there is no evidence of differential predictive validity for the MAPS subscales across ethnoracial identities predicting child psychopathology.

Discussion

The goals of this study were to (a) refine the MAPS dimensions, items, and item formats using parents' feedback from qualitative cognitive interviews; (b) establish the factor structure of the Revised MAPS in a racially and ethnically diverse national sample using confirmatory factor analyses; and (c) establish the reliability, validity, and MI of the Revised MAPS in a racially and ethnically diverse sample. Using a two-phase mixed-methods approach, the Revised MAPS was developed, its factor structure was confirmed, and strong psychometric properties were established.

Qualitative Interviews

The first goal of this study was to refine the MAPS using qualitative cognitive interviews guided by previous work (DeWalt et al., 2007; Willis, 2004). The refinement of the MAPS through qualitative cognitive interviews underscored the critical importance of linguistic precision and cultural relevance in the assessment of parenting. By meticulously revising items based on feedback from parents, this study not only addressed ambiguities in wording, such as the use of terms like "argue" and "lose my temper," but also tackled the issue of how cultural perspectives influence the interpretation of parenting. For instance, the revision prompted by concerns over assumptions of children having their own rooms highlights the subtle ways in which

Table 5
Associations Among, and Descriptive Statistics for, Key Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. MAPS H	—												
2. MAPS L	.596**	—											
3. MAPS P	.667**	.545**	—										
4. MAPS PP	.064**	−.034	−.027	—									
5. MAPS R	−.126**	−.089**	−.212**	.677**	—								
6. MAPS S	−.205**	−.130**	−.250**	.616**	.737**	—							
7. MAPS W	−.168**	−.073**	−.156**	.510**	.612**	.669**	—						
8. PHQ-9	.457**	.383**	.431**	−.019	−.136**	−.156**	−.162**	—					
9. GAD-7	.352**	.267**	.339**	.064**	−.013	−.047	−.070**	.767**	—				
10. APQ PP	−.131**	−.092**	−.164**	.438**	.577**	.513**	.476**	−.113**	−.030	—			
11. APQ ID	.534**	.694**	.490**	−.039	−.143**	−.186**	−.103**	.380**	.243**	−.045	—		
12. APQ PS	.490**	.547**	.537**	−.217**	−.344**	−.350**	−.317**	.399**	.253**	−.232**	.571**	—	
13. Child psy.	.539**	.506**	.566**	−.137**	−.287**	−.309**	−.247**	.623**	.501**	−.221**	.452**	.538**	—
M (*SD*)	15.54 (5.82)	15.82 (5.86)	7.08 (3.82)	22.73 (4.86)	16.69 (3.38)	12.87 (2.33)	17.02 (3.17)	6.94 (6.55)	2.04 (3.49)	13.40 (2.30)	7.68 (2.96)	5.73 (3.29)	23.01 (6.79)

Note. MAPS = Multidimensional Assessment of Parenting Scale; MAPS H = MAPS hostility; MAPS L = lax control; MAPS P = physical control; MAPS PP = proactive parenting; MAPS R = reinforcement; MAPS S = supportiveness; MAPS W = warmth; PHQ-9 = Patient Health Questionnaire–9; GAD-7 = General Anxiety Disorder–7; APQ PP = Alabama Parenting Questionnaire positive parenting; APQ ID = APQ inconsistent discipline; APQ PS = APQ poor supervision; Child psy. = child psychopathology.
** $p = 0.01$ (two-tailed).

socioeconomic and cultural contexts shape our understanding of parenting norms and practices. The incorporation of both nonphysical and verbal forms of affection into the MAPS reflects an appreciation for the diverse ways in which warmth can be expressed across cultures. This adjustment, rooted in cognitive interview findings, directly responds to the previously noted gap in capturing the full spectrum of affectionate behaviors in parenting scales. This evolution of the MAPS items demonstrates a commitment to inclusivity, recognizing that expressions of warmth and discipline are culturally bound and can vary significantly across different familial and cultural backgrounds.

Further, issues remain in terms of item wording—for example, regarding the cultural nuances not fully captured in the term “chores” and its implications for diverse racial and ethnic groups, particularly Asian American and Latinx families. As such, our discussion would be incomplete without acknowledging the complexities surrounding language and its cultural relevance in research instruments. While our methods adhered to established *a priori* thresholds for item revision, quantitative thresholds alone may not fully encompass the subtleties of cultural differences in parenting practices. Our strict adherence may represent a significant gap in our approach, underscoring the need for future research to further explore, understand, and integrate cultural contexts into the development and refinement of parenting assessments. While the absence of significant DIF in our results provided some reassurance against systematic bias, it does not negate the potential for cultural insensitivity in item wording or content. Therefore, we recognize the importance of continually striving for measurement tools that are not only psychometrically robust but also culturally sensitive and inclusive. This study therefore serves to highlight the critical need for better methods of item revision and validation. Future research should aim to bridge the gap between statistical validity and cultural relevance in the development of tools like the MAPS, and particularly in the assessment of parenting more generally.

This study benefited from starting with items that had undergone extensive empirical research (Parent & Forehand, 2017; Rodriguez, Cadet, et al., 2023) but had not been subjected to formal cognitive interviewing with its target population. Subjecting survey items to cognitive interviewing is a standard survey development technique, particularly for large-scale questionnaires (Johnson, 2016). The use of cognitive interviews, however, highlighted the pivotal role of cultural considerations in the development of parenting scales and underscored the limitations of relying solely on quantitative methods to understand the nature of parenting in diverse populations. By integrating findings from cognitive interviews, this study not only enhances the cultural sensitivity of the MAPS but also sets a precedent for future research to further explore the cultural contexts that shape parenting, building on the utility of these methods in other disciplines (Nicolaidis et al., 2015). This approach aligns with calls for more multimethod research strategies, particularly in fields like developmental psychology, which has been critiqued for its overreliance on Western, educated, industrialized, rich, and democratic populations (Nielsen et al., 2017). Through this lens, the study contributes to a growing body of literature advocating for the incorporation of diverse perspectives in research, ultimately aiming to create more inclusive and representative measures of parenting (Rodriguez, La Barrie, et al., 2023).

MAPS Factor Structure

The second goal of the present study was to establish the factor structure of the Revised MAPS in a national U.S. sample, oversampling for racially and ethnically minoritized groups. Unlike the original validation study (Parent & Forehand, 2017), our findings through ESEM and subsequent CFA indicated a shift toward a more parsimonious factor structure. This resulted in the CFA model being the chosen model due to both acceptable to excellent fit to data and better measurement quality indices, as well as being the most parsimonious model. Statistically speaking, the chosen model suggested that simply correlating the distinct factors resulted in acceptable model fit indices as well as better measurement quality indices, such that the Revised MAPS assesses seven robust, distinct, yet correlated dimensions of parenting. Contrary to the original MAPS validation study, hierarchical factor structures, both including second- and third-order factors accounting for Negative and Positive Broadband subscales as well as a "Parenting" third-order factor, did not provide a good fit to the data when using both traditional CFA and ESEM to estimate these models. This was not surprising, given that as noted by Parent and Forehand (2017), this hierarchical conceptualization of parenting was a novel conceptualization of parenting largely driven by the better fit of hierarchical factor analyses over other models. However, a hierarchical structure is not theoretically supported by Schaefer's (1959) circumplex model of maternal behavior, which guided the construction of the MAPS. Despite these differences from the original MAPS validation study, the factor structure of the MAPS subscales was strongly supported in this diverse sample.

In summary, the factor structure differences elucidated in this study contribute to a broader discourse on the need for culturally attuned, empirically validated measures within the field of developmental psychology. As we move toward a more inclusive understanding of parenting, it becomes imperative that our assessment tools evolve in tandem to capture the full spectrum of parenting behaviors across different cultural backgrounds, and diverse family structures (e.g., Rodriguez et al., 2024). Our findings not only validate the Revised MAPS as a significant step in this direction but also call for continued exploration and adaptation of parenting scales to better serve diverse populations.

Psychometric Properties: Reliability, Validity, and MI

The third goal of this study was to establish the reliability, validity, and MI of the Revised MAPS in a racially and ethnically diverse sample. The reliability of the MAPS was strongly supported. In terms of validity, the Revised MAPS demonstrated partial convergent and predictive validity, particularly highlighted by the negative correlations between positive parenting dimensions (e.g., warmth, supportiveness) and child psychopathology. These findings are in line with a wealth of literature suggesting that positive parenting is inversely related to adverse child psychopathology (Rose et al., 2018), thereby reaffirming the theoretical underpinnings of the MAPS constructs. However, it is worth noting that there were smaller magnitude correlations with parental anxiety, and this was also the case for the APQ. Inconsistent discipline was strongly correlated with MAPS lax parenting, positive parenting with reinforcement, proactive parenting, supportiveness, and warmth. Other correlations were smaller in magnitude (e.g., APQ inconsistent discipline and MAPS proactive parenting). Smaller magnitude correlations may signal the need for further exploration and understanding of these associations. It is important to also note that these associations were observed in a racially and ethnically diverse sample, while the APQ has been found to lack MI by ethnoracial identity (Shaffer et al., 2022).

In validating a scale in heterogenous samples, MI is an essential psychometric property that refers to the stability of scale across groups, which could be demographic groups, or meaning, across time. Smaller sample sizes, or even those that are census-matched and do not recruit large samples of racially and ethnically minoritized groups, may underestimate MI. For example, while a few studies have tested MI in parenting scales in the last couple of years using large samples, but these studies were limited by uneven group sizes (Rodriguez, Cadet, et al., 2023; Shaffer et al., 2022). As such, enough participants from minoritized racial and ethnic groups needed to be oversampled to be able to test for MI. In this study, IRT DIF item-level analyses were used to test for MI given that this method provides clearer guidance for removal than factor analytic approaches. Although initial tests flagged several items for DIF, pseudo-R^2 did not meet the recommended thresholds, which tend to be much higher (e.g., Zumbo, 1999). More recent studies have suggested that these thresholds are too large and potentially result in underidentification of DIF. A potential solution to this problem was the use of Monte Carlo simulations proposed by Choi et al. (2011), as implemented by the R package *lordif*.

Notably, previous research identified DIF by racial and ethnic groups in several items related to negative parenting (Rodriguez, Cadet, et al., 2023; Rodriguez et al., 2024). The cognitive interviews led to revisions of these items to reduce ambiguity and confusion, addressing potential sources of DIF. This meticulous revision process underscores the dynamic nature of MI and the importance of continuous evaluation to ensure the scale's broad applicability. The revisions made to items such as "I am afraid that disciplining my child for misbehavior will cause her/him to not like me," among others, highlight the critical role of cognitive interviews in enhancing the scale's clarity and cultural sensitivity (Willis, 2018). These efforts to refine the MAPS through cognitive interviews not only aim to mitigate previously identified DIF but also reflect an ongoing commitment to ensuring the scale's relevance and utility across diverse populations. This study's findings call for further research to examine how various factors influence parenting behaviors and child outcomes, beyond the ones assessed in this study, aiming to refine the MAPS into a more culturally sensitive and psychometrically robust tool. The insights gained from this process not only contribute to the existing body of literature on parenting assessment but also underscore the necessity of incorporating diverse perspectives in the development and evaluation of such measures.

Ethnoracial Comparisons in Revised MAPS Subscales

With greater confidence in ethnoracial comparisons in MAPS subscales after testing for MI and confirming that MI was achieved, MAPS subscales were compared across racial and ethnic groups. No mean differences emerged in Lax Control, Proactive Parenting, Supportiveness, and Warmth as part of the Revised MAPS subscales. However, mean differences emerged for Hostility, Physical Control, and Reinforcement—specifically, non-Hispanic White parents reported higher levels of Hostility, Physical Control, and Reinforcement than non-Hispanic Black and Hispanic parents, even after adjusting for multiple comparisons and covariates. Furthermore, both non-Hispanic White and Black parents, as well as Hispanic parents, exhibited higher Physical Control compared to non-Hispanic Asian parents. The lower means of Hostility and Physical Control among non-Hispanic Black and Hispanic parents relative to non-Hispanic White parents are contrary to studies reporting higher levels of different forms of physical punishment among non-Hispanic Black and Hispanic parents compared with non-Hispanic White parents, including national epidemiological surveys (Finkelhor et al., 2019; Taillieu et al., 2014). Findings regarding Hispanic families have generally been mixed, with epidemiological trends suggesting an increase in physical punishment over time (Taillieu et al., 2014) and others suggesting Hispanic parents are less likely to use physical punishment compared with non-Hispanic Black parents (Finkelhor et al., 2019). Previous work has also highlighted psychometric issues, including a lack of MI, in the measurement of physical punishment and negative parenting in racially and ethnically minoritized people and parents (Rodriguez et al., 2019; Rodriguez, Cadet, et al., 2023), which may contribute to the mixed findings in the literature regarding physical punishment.

Considering that the findings from the present study differ from prior literature, the lower levels of physical punishment among non-Hispanic Black and Hispanic parents relative to non-Hispanic White parents may be explained by qualitative findings from this study. For example, a Black father expressed that "exploding in anger" (originally Item 13) may be interpreted as *physical violence* when reported by a Black father, but as an *emotional* outburst when reported by a White parent. This may reflect well-documented patterns in which racially and ethnically minoritized parents are disproportionately likely to suffer negative consequences when using or disclosing physical punishment. Lastly, it must be highlighted that previous studies showing higher levels of physical punishment among ethnoracially minoritized parents did not use validated measures to assess physical punishment (e.g., Finkelhor et al., 2019; Taillieu et al., 2014). As such, the ethnoracial group differences emerging from this study benefited from the current dedicated psychometric evaluation addressing ethnoracial biases in the MAPS as well as preliminary studies documenting measurement issues related to addressing psychometric issues in measuring negative parenting, including physical control (Rodriguez, Cadet, et al., 2023). In fact, in recent times, few studies have addressed MI by ethnic and racial identities (Rodriguez, La Barrie, et al., 2023). As such, the variations between current findings and previous research may be a result of the revisions intended to address ethnoracial biases in the measurement of parenting and further underscore the need for measures that do not mischaracterize specific groups.

More broadly, although these mean comparisons may be informative in understanding potential differences in parenting among parents across racial and ethnic groups, individuals from racially and ethnically minoritized groups are not monoliths. As such, further research is needed on the *unique* experiences of individuals from racially and ethnically minoritized groups with regard to parenting, including negative parenting practices. While this study's primary interest lay in capturing aspects of parenting that might be most universal for use in diverse context, the qualitative findings suggest that ethnoracial identity-specific factors may also play a critical role, warranting further exploration.

Limitations, Strengths, and Future Directions

Research on the MAPS has pointed to a lack of MI by race and ethnicity as it relates to *negative* parenting but not *positive* parenting (Rodriguez, Cadet, et al., 2023), a finding that was replicated in this study; a lack of MI in the APQ has been found in items related to positive *and* negative parenting (Shaffer et al., 2022). The APQ items, along with items from other scales, were modified in developing the MAPS. This finding suggests that the measurement of parenting has improved in a few psychometric studies. Though this study contributed to these efforts through the enhancement of the psychometric properties of the MAPS for diverse populations, it is not without limitations. This study was based on a self-report of parent and child behavior and is cross-sectional, and participants were recruited from QualtricsXM panels. Participants in both phases were more educated and had higher incomes than participants in clinical, community, and online samples in previous MAPS research (Rodriguez, Cadet, et al., 2023). Possible future directions could involve the integration of coparent or child reports, direct behavioral observations, or ecological momentary assessments. Such multi-method approaches could address the concerns related to shared method variance but also enrich our understanding of the constructs measured by the MAPS. The qualitative phase of this study was limited by interviews conducted via Zoom. Future research could examine whether internet-based versus in-person interviews provide similar or different information (Willis, 2018). However, internet-based recruitment facilitated engaging racially and ethnically diverse parents, including fathers, who are underrepresented in research. Across both phases of the study, there are limitations associated with using race and ethnicity alone. Future studies may consider measuring variables related to social position to increase the representation of other underrepresented groups and consider intersectionality. Nevertheless, given that recent movements to increase diversity have focused on racial and ethnic representation, it is important to consider race and ethnicity in psychometric research, such that individuals from minoritized racial and ethnic groups are not mischaracterized.

Notwithstanding these limitations, this study advances the field by not only using a mixed-methods approach to enhance the psychometric properties of MAPS but also validating the MAPS in a racially and ethnically diverse sample. Importantly, this included *fathers from* racially and ethnically minoritized groups representing across the United States. Though the field has made progress in increasing the representation of fathers in psychometric parenting research (e.g., Rodriguez, Cadet, et al., 2023; Shaffer et al., 2022), this has not resulted in increasing the representation of fathers from minoritized racial and ethnic backgrounds. Similarly, systematic reviews in the past few decades on the psychometric properties of parenting scales have pointed to the scarcity of psychometric research on parenting measures (Hurley et al., 2014; Locke & Prinz, 2002; Rodriguez et al., 2024). According to a more recent systematic review, the current state of the field of psychometric evidence of parenting scales on parents from racially and ethnically minoritized backgrounds was not much more hopeful (Rodriguez, La Barrie, et al., 2023). Therefore, this study provides an example of how the communities we study can make important contributions and how researchers can facilitate an empowering and power-sharing process that attends to a more inclusive measurement of parenting. In that sense, this study provides examples of ways to include parents from minoritized racial and ethnic backgrounds. In addition, given that this study was made possible by starting with an already strong and theoretically driven measure, we capitalized on decades of research, theoretical advances, and analytic developments across several decades.

References

Bornstein, M. H. (2013). Parenting and child mental health: A cross-cultural perspective. *World Psychiatry*, *12*(3), 258–265. https://doi.org/10.1002/wps.20071

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Press.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, *39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

DeWalt, D. A., Rothrock, N., Yount, S., Stone, A. A., & the Patient-Reported Outcomes Measurement Information System Cooperative Group. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, *45*(5), S12–S21. https://doi.org/10.1097/01.mlr.0000254567.79743.e2

Finkelhor, D., Turner, H., Wormuth, B. K., Vanderminden, J., & Hamby, S. (2019). Corporal punishment: Current rates from a national survey. *Journal of Child and Family Studies*, *28*(7), 1991–1997. https://doi.org/10.1007/s10826-019-01426-4

Frick, P. J. (1991). *The Alabama Parenting Questionnaire* [Unpublished Rating Scale]. University of Alabama.

González-Romá, V., Hernández, A., & Gómez-Benito, J. (2006). Power and Type I error of the mean and covariance structure analysis model for detecting differential item functioning in graded response items. *Multivariate Behavioral Research*, *41*(1), 29–53. https://doi.org/10.1207/s15327906mbr4101_3

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Child Psychology & Psychiatry & Allied Disciplines*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

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

Johnson, C. (2016). *Proposal guidelines for new content on the 2007–2008 National Health and Nutrition Examination Survey (NHANES)*. Retrieved December 13, 2006, from https://wwwn.cdc.gov/nchs/nhanes/proposalguidelines.aspx

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1–3), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Lindhiem, O., & Shaffer, A. (2017). Introduction to the special series: Current directions for measuring parenting constructs to inform prevention science. *Prevention Science*, *18*(3), 253–256. https://doi.org/10.1007/s11121-016-0724-6

Locke, L. M., & Prinz, R. J. (2002). Measurement of parental discipline and nurturance. *Clinical Psychology Review*, *22*(6), 895–929. https://doi.org/10.1016/S0272-7358(02)00133-2

Matza, L. S., Murray, L. T., Phillips, G. A., Konechnik, T. J., Dennehy, E. B., Bush, E. N., & Revicki, D. A. (2015). Qualitative research on fatigue associated with depression: Content validity of the Fatigue Associated with Depression Questionnaire (FAsD-V2). *The Patient: Patient-Centered Outcomes Research*, *8*(5), 433–443. https://doi.org/10.1007/s40271-014-0107-7

Nicolaidis, C., Raymaker, D., Katz, M., Oschwald, M., Goe, R., Leotti, S., Grantham, L., Plourde, E., Salomon, J., Hughes, R. B., Powers, L. E., & the Partnering with People with Disabilities to Address Violence Consortium. (2015). Community-based participatory research to adapt health measures for use by people with developmental disabilities. *Progress in Community Health Partnerships*, *9*(2), 157–170. https://doi.org/10.1353/cpr.2015.0037

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38. https://doi.org/10.1016/j.jecp.2017.04.017

Parent, J., & Forehand, R. (2017). The Multidimensional Assessment of Parenting Scale (MAPS): Development and psychometric properties. *Journal of Child and Family Studies*, *26*(8), 2136–2151. https://doi.org/10.1007/s10826-017-0741-5

Rodriguez, V. J., Cadet, G. J., Sisitsky, M., Cooley, C., Acosta, J., Coles, E., Charity-Parker, B., Walters, A., Shaffer, A., & Parent, J. (2023). Assessing parenting in racially and ethnically diverse families: A lack of measurement equivalence. *Journal of Family Psychology*, *37*(6), 753–762. https://doi.org/10.1037/fam0001095

Rodriguez, V. J., La Barrie, D. L., Zegarac, M., & Shaffer, A. (2023). A systematic review of parenting scales the literature on measurement invariance/equivalence of parenting scales by race and ethnicity: Recommendations for inclusive parenting research. *Assessment*, *30*(1), 22–36. https://doi.org/10.1177/10731911211038630

Rodriguez, V. J., LaBarrie, D. L., Adams, S. J., & Liu, Q. (2024). Parenting measures and their psychometrics in LGBTQIA families: A systematic review. *Child Psychiatry & Human Development*. Advance online publication. https://doi.org/10.1007/s10578-024-01672-8

Rodriguez, V. J., Shaffer, A., Are, F., Madden, A., Jones, D. L., & Kumar, M. (2019). Identification of differential item functioning by race and ethnicity in the Childhood Trauma Questionnaire. *Child Abuse & Neglect*, *94*, Article 104030. https://doi.org/10.1016/j.chiabu.2019.104030

Rose, J., Roman, N., Mwaba, K., & Ismail, K. (2018). The relationship between parenting and internalizing behaviours of children: A systematic review. *Early Child Development and Care*, *188*(10), 1468–1486. https://doi.org/10.1080/03004430.2016.1269762

Schaefer, E. S. (1959). A circumplex model for maternal behavior. *The Journal of Abnormal and Social Psychology*, *59*(2), 226–235. https://doi.org/10.1037/h0041114

Shaffer, A., Rodriguez, V. J., Kolko, D. J., Pilkonis, P. A., & Lindhiem, O. (2022). Measurement invariance in the Alabama Parenting Questionnaire: Implications for culturally inclusive parenting research. *Journal of Family Issues*, *43*(8), 2040–2062. https://doi.org/10.1177/0192513X211030022

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Taillieu, T. L., Afifi, T. O., Mota, N., Keyes, K. M., & Sareen, J. (2014). Age, sex, and racial differences in harsh physical punishment: Results from a nationally representative United States sample. *Child Abuse & Neglect*, *38*(12), 1885–1894. https://doi.org/10.1016/j.chiabu.2014.10.020

United States Census Bureau. (2019). *QuickFacts: United States*. https://www.census.gov/quickfacts/fact/table/US/PST045219

van Zyl, L. E., & Ten Klooster, P. M. (2022). Exploratory structural equation modeling: Practical guidelines and tutorial with a convenient online tool for Mplus. *Frontiers in Psychiatry*, *12*, Article 795672. https://doi.org/10.3389/fpsyt.2021.795672

Willis, G. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.

Willis, G. (2018). Cognitive interviewing in survey design: State of the science and future directions. In D. Vannette, & J. Krosnick (Eds.), *The Palgrave handbook of survey research* (pp. 103–107). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-54395-6_14

Zumbo, B. D. (1999). *A handbook on the theory and methods of Differential Item Functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense.

Received December 12, 2023
Revision received August 20, 2024
Accepted August 25, 2024 ■