# Temperament in Middle Childhood Questionnaire: New data on factor structure and applicability in a child clinical sample

**Michael Kozlowski**[1], **Dylan Antovich**[1], **Sarah Karalunas**[2], **Joel Nigg**[1]

[1]Department of Psychiatry, Oregon Health and Science University

[2]Department of Psychological Sciences, Purdue University

## Abstract

The Temperament in Middle Childhood Questionnaire (TMCQ) is one of a family of instruments representing one of the major conceptual models of child temperament. The present study reports new psychometric information on the TMCQ using a larger sample than in prior factor analytic studies of this instrument. Data from parent ratings of 1,418 children were utilized. The sample of community volunteers included 697 typically developing youth and 721 defined by research diagnostic procedures as having attention-deficit-hyperactivity disorder (ADHD). Results failed to support the original proposed structure of the TMCQ, but found support for a structure with 12 subscales that confirmed a substantial portion of the lower order factor structure. However, the intended 3-factor higher order structure was not able to be fully recovered. Two-group invariance was supported in the final model, supporting use in studies of typical and atypical development. In conclusion, with some modifications the TMCQ remains a useful research measure at the lower order factor level. The validity of the higher order structure is less clear, likely due to measure-specific limitations, and suggests a need for some refinement to the measure.

## Keywords

Questionnaire-based assessment of child temperament has emerged as a cost effective yet powerful tool in developmental, neuroscience, and psychopathology research as a complement to observational measures. Leading conceptual models and methods have grown to parallel the questionnaire assessment of adult personality. However, child temperament questionnaires are sometimes less well-developed in terms of both content validity and clarity of factor structure than comparable adult or child personality scale measures (Shiner, Soto, & De Fruyt, 2021).

One of the most venerable and theoretically sophisticated sets of ratings-based child temperament measures has derived from the laboratory of Mary Rothbart and colleagues. Her team developed a series of measures designed for different developmental periods: the Infant Behavior Questionnaire (IBQ; Rothbart, 1981), Child Behavior Questionnaire

Correspondence concerning this article, including requests for data-sharing and corresponding codes, should be addressed to Joel T. Nigg, 3161 SW Pavilion Loop, Mailcode: UHN 80R1, Portland, Oregon 97239-3098. niggj@ohsu.edu.

(CBQ; Rothbart, Ahadi, Hershey, & Fisher, 2001), Temperament in Middle Childhood Questionnaire (TMCQ; Simonds & Rothbart, 2004), and Early Adolescent Temperament Questionnaire (EATQ; Capaldi and Rothbart, 1992). These measures and the model they are based on are widely used and considered one of the major contemporary models of temperament (Mervielde & De Pauw, 2012). They followed a sophisticated developmental theory and were developed in a rational process, with conceptually relevant items created and tested, factor analytic evaluation in small demonstration samples, and model-based organization (Rothbart, 2011). All were intended to reflect the same underlying regulation-reactivity model of temperament with three or four higher order factors, but varying numbers of lower order scales. The set of measures and the theory and model behind them also enjoy rich linkage to animal and neuroscience research (Posner & Rothbart, 2018), adding to their value for inferences ranging from basic to translational studies of development and psychopathology.

However, concerns have been raised about the psychometric properties of these measures, particularly those designed to assess temperament in middle childhood such as the TMCQ (Kotelnikova, Olino, Klein, Mackrell, & Hayden, 2017; Nystrom & Bengtsson, 2017). One issue has been the lack of formal measurement modeling for construct validity during the scale development process. Another is that previous investigations relied on small sample sizes to validate the model, rendering it difficult to be confident of the reproducibility of scale reliability as well as lower and higher order factor structure in larger samples.

## The TMCQ

Our focus on the TMCQ in particular, due to its importance for studies of development and psychopathology in middle childhood, the peak years of onset for common psychopathology and a critical period for child development (Shiner, 2021), and the shortage of validation studies of it. The TMCQ has been used for multiple studies of cognitive, personality, and neural development (Affrunti, Geronimi, & Woodruff-Borden, 2014; Affrunti & Woodruff-Borden, 2015; Ato, Fernández-Vilar, & Galián, 2020; Inuggi et al., 2014) as well as developmental psychopathology (Karalunas, Gustafsson, Fair, Musser, & Nigg, 2019; Kotelnikova, Mackrell, Jordan, & Hayden, 2015; Kotelnikova et al., 2017; Nigg et al., 2020; Rutter & Arnett, 2020). The TMCQ is designed to evaluate temperament in the regulation-reactivity model from ages 7–10, although there is evidence that such questionnaires can be used and are used outside their originally intended age ranges (Soto & John, 2014) and can be completed by adult raters or by self-report (Simonds & Rothbart, 2004, 2009). It was initially described as containing 17 scales capturing lower order developmental traits, of which 13 were derived quite closely from the original CBQ and four were added to reflect the growing capabilities of developing children in this age range along with new items written specifically to capture temperament traits in middle childhood as well as items adapted from the Hampton Individual Differences Questionnaire (Baker & Victor, 2001), Childhood Temperament and Personality Questionnaire (CTPQ; Victor, Rothbart, & Baker (2003) and the Berkeley Puppet Interview self-report version of the CBQ (CBQ-BPI; Ablow & Measelle (1993). With regard to its psychometric validity, Simonds and Rothbart (2004) evaluated 95 typically developing children by parent report, and reported satisfactory scale reliability data for 15 of the scales. Simonds (2006) noted that only 141 items were retained

to achieve scale reliability, and Simonds and Rothbart (2009) suggested that the *Activation Control* scale is experimental. All of this left an option for 14, 15, or 17 subscales. It was hypothesized by those authors that the subscales form a superordinate structure of three higher order factors: Surgency, Negative Affectivity, and Effortful Control. A fourth factor for Affiliativeness or Affiliation was also suggested, but never formally investigated (Simonds & Rothbart, 2004).

Nystrom and Bengtsson (2017) undertook an analysis to test the hypothesis of three superordinate factors, in 157 typically developing children aged 7–11 from Sweden. They accepted at face value the recommended 14 subscales (Simonds & Rothbart, 2009) and attempted to fit them in a CFA to a 3-factor solution (Effortful Control, Surgency, Negative Affect). Although their results did not support a completely distinct factor solution, allowing scale cross-loadings onto higher order factors, yielded a generally acceptable three-factor solution. The solution included cross-loadings from Activating Control, Attentional Focusing, and Inhibitory Control onto Negative Affect; a cross-loading from Impulsivity onto Effortful Control; and cross-loadings from Attentional Focusing and Inhibitory Control onto Surgency.

More recently, Lipska et al. (2021) undertook a similar effort to that of Nystrom and Bengtsson (2017) with responses collected from parent ratings of 189 children aged 7–10 in Poland. Using a dimension reduction procedure from network psychometrics known as exploratory graph analysis (Golino et al., 2020), they identified a modified solution with 16 scales that yielded a three-factor higher order solution in 68 out of 100 bootstrapped solutions (with Shyness excluded). Subscales represented theorized higher order factors in some iterations of the model. Across all iterations, however, support for the theorized higher order structure was dimensionally instable: Effortful Control was the most reliably represented by Low Intensity Pleasure and Perceptual Sensitivity; Negative Affectivity by Anger, Sadness, and Reversed Soothability along with Fear and Discomfort in 98 out of 100 solutions. Constructs traditionally understood to load on Surgency loaded in some resamples, but no clear pattern of loadings emerged across all solutions. They also compared their EGA findings with a traditional parallel analysis, which supported a three-factor solution.

The only previous item-level factor analysis of the TMCQ was reported by Kotelnikova et al. (2017) in a convenience sample of 654 typically-developing 9-year-old children in Canada and the United States. Their results suggested that 92 of the items loaded on 13 interpretable factors. However, they concluded that many of their identified factors did not resemble one of the original scales proposed. They also investigated the proposed higher order factor structure and were unable to confirm the hypothesized 3 factor solution using their 13 factors; proposing instead a three-factor solution of Impulsivity/Negative Affectivity, Negative Affectivity, and Openness/Assertiveness based on their new 13 scales. Importantly, Effortful control and Extraversion/Surgency were not clearly represented. The striking contrast between their results and others suggested a significant ongoing need for further examination and evaluation of the item-level and higher order structure of the instrument.

### What is the Factor Structure of the TMCQ?

The aforementioned studies raise doubts about the construct validity and appropriate factor structure of TMCQ as initially proposed, yet this literature also suggests that some elements of the instrument produce robust scales. While it could be that the TMCQ requires revision to achieve its theoretical intent, only a limited number of factor structures have been explored in previous research. Rothbart (2011) conceptual model of temperament implies a higher order structure where the items are to load onto first order latent variables, which in turn load on higher order latent variables, but this hypothesis also has never been formally tested. Another caveat is that the final models in previous research included cross-loadings on the first or second order factor solutions. Given the substantial length of the TMCQ as well as the number of first and second order factors and the conceptual overlap of the scales in relation to the hypothesized higher order factors, this is unsurprising. For instance, two prior psychometric investigations (Kotelnikova et al., 2017; Nystrom & Bengtsson, 2017) found a strong association between items expected to load on distinct factors of anger and sadness, yet both scales are hypothesized as components of negative affectivity either directly or via correlated residuals. Thus, like many personality measures (such as the Big Five; Marsh et al. (2010), the factor structure of the TMCQ may be substantially complex perhaps requiring a robust examination of the factor structure from multiple modeling perspectives.

## Current Study

The purpose of the current study then was to conduct a comprehensive re-evaluation of item-level and higher-order factor structures of the TMCQ, to refine a factor structure that is largely guided by theory, and to further extend information on the factor structure to a child psychopathology sample. Like Kotelnikova et al. (2017), we used a convenience sample of typically developing, community recruited volunteers (albeit extensively screened for normal range IQ and psychopathology status and entirely from the United States). We systematically evaluated a series of models. We first tested competing models to attempt to replicate some prior proposals; then determined the best solution in the present data set. Novel here are (1) the inclusion of a similar sized sample, from the same community of children, with attention-deficit / hyperactivity disorder (ADHD) to determine generalizability in an atypically developing and commonly studied sample and (2) the application of a modeling framework that explicitly accounts for a substantially complex theoretical structure.

## Methods

### Participants

The sample included 1,418 children ranging in age from 6.9 to 13.8 years of age (81% of the sample was aged 7–10 years. We also report sensitivity analyses in the supplementary material using strict age parameters, but in view of the fact that such instruments tend to be sometimes used outside the theorized age range, we proceeded to report on the full sample.) The youth were part of a larger study of development (Karalunas et al., 2014; Nigg et al., 2020). The sample used here comprised 721 children with ADHD and 697

typically developing children without ADHD; in both groups other psychopathology was free to vary with limited exclusions noted below. Thus, it provided both (a) a reasonable sized sample of typically developing children similar to Kotelnikova et al. (2017) and (b) a similar sized sample of children with a common developmental psychopathology for which relations to temperament are of considerable interest. Tests of multiple group invariance justified reporting full sample solutions. Overall, the sample provides both the largest factor analyzed sample on the TMCQ to date, and data on generalizability across typical and atypical development.

### Sample Ascertainment and Description

Data were collected at the baseline of an ongoing longitudinal developmental study. This study was approved by institutional IRB. Volunteers were recruited from the community via mass mailings, using commercial mailing lists, to all families with children in the target age range within the geographic radius of 50 miles from a university located in the Pacific Northwest. One set of mailings requested volunteers with children with possible or definite ADHD; a second set for those with generally healthy, typically developing children with no history of identified learning or attention problems. In response to mailings to parents of all children in the target age range in the intended catchment area, we received $n = 2144$ inquiries (Based on number of mailings and estimated base rate of ADHD of 5%, this yielded an estimated response rate of about 1% for non-ADHD participants and about 30% for ADHD participants). An initial screening phone call served to establish eligibility and interest. Nearly half were ruled out at this stage due to long acting psychiatric medications, exclusionary medical conditions, or lack of interest. Those excluded did not differ reliably from the final sample on sex ratio ($p = .11$) or race ($p = .22$), but were marginally lower income ($p = .06$) and slightly younger ($p = .06$).

The remaining participants ($n = 1449$) completed extensive structured interview and multi-informant questionnaire data to ascertain ADHD and non-ADHD assignments, other psychopathology, and eligibility for other studies conducted with this cohort. A primary caregiver completed the TMCQ at this visit. Children were excluded at this second stage if the assessment identified a history of non-febrile seizure, head injury with loss of consciousness (> 60 seconds), diagnosis of autism spectrum disorder or intellectual disability, or other major medical conditions that had been missed at the initial screen. This resulted in the final $n$ of 1418. Table 1 provides description of the sample as well as of the two subgroups.

### Temperament in Middle Childhood Questionnaire (TMCQ)

As noted, the TMCQ was developed by Simonds and Rothbart (2004) as an extension of the Child Behavior Checklist (CBQ), designed for young children, to now assess children aged 7–10 years. The parent rated form was studied here. It included 157 items rated on a 5-point Likert scale and designed to assess 17 traits: Activation Control, Activity, Affiliativeness, Anger, Fear, High Intensity Pleasure [HIP], Impulsivity, Inhibition, Sadness, Shyness, Soothability, Assertiveness, Attentional Focus, Low Intensity Pleasure [LIP], Perceptiveness, Discomfort, and Openness. Fourteen of these 17 lower order scales (13 derived from the CBQ plus activation control) were combined to produce three higher order factors: Effortful

Control, Negative Affectivity, and Surgency; Simonds & Rothbart, 2009). However, in view of the goal of clarifying factor structure and following the only prior item-level analysis of this instrument (Kotelnikova et al., 2017), all 157 items and all 17 scales were included in the current report.

## Analysis Plan

The plan for analysis was to establish construct validity by subjecting the data to the hypothesized 17 scale TMCQ structure initially proposed by Rothbart and colleagues (Simonds & Rothbart, 2009). While we planned to investigated alternative solutions using exploratory (EFA) and confirmatory factor analysis (CFA) as needed, our principal method was exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009). ESEM is well-suited to the expectation of cross-loaded models by representing complex factor structures by combining features of EFA and CFA models while evaluating models in a full structure equation framework (Marsh, Morin, Parker, & Kaur, 2014). ESEM can estimate factors a priori, like a traditional CFA, but allows for indicators to load on all latent variables in a rotated solution. The advantage to such an approach is the ability to model solutions where well-defined simple structures may rely on small, but meaningful cross-loadings among the factors, which the prior TMCQ literature already consistently indicated is necessary. Indeed, many self-report measures contain distinct yet interrelated constructs (Marsh et al., 2014). ESEM has been utilized successfully to represent complex factor structures in personality (Marsh et al., 2010; Ng, Cao, Marsh, Tay, & Seligman, 2017) and developmental research (Marsh, Nagengast, & Morin, 2013).

Analyses for factor analytic models were conducted in Mplus version 8.7 (Muthén & Muthén, 1998–2017) using the complex command to account for family-level clustering (that is, 239 sibling pairs in the sample). Missingness on item level TMCQ data (< 5% of item level responses) was handled with full information maximum likelihood. Models were evaluated using the full sample as well as both the ADHD and typically developing subsamples separately. For all ESEM models, a geomin rotation was employed (Morin, Marsh, & Nagengast, 2013) with 100 random starts (Hattori, Zhang, & Preacher, 2017). Determination of fit relied on combination rules previously established via simulation studies. These were: the comparative fit index (CFI) 0.90, Tucker Lewis index (TLI) 0.90, root mean square error of approximation (RMSEA) 0.08, but preferably 0.05, and standardized root mean-square residual (SRMR) < 0.08 (Hu & Bentler, 1999). We included the chi-square model test, but because significance level depends on sample size, models were not evaluated solely based on it. For all models, correlated errors (residuals) were allowed on a limited basis: for items loading on the same scale and with similar item wording and content based on the consensus among the authors (e.g., "Tends to say the first thing that comes to mind, without stopping to think about it" (item 16); "Says the first thing that comes to mind" (item 25). Workflow occurred in three phases as described next.

**Phase 1: Tests of A Priori or Previously Proposed Factor Structures—**We began by thoroughly evaluating prior models using CFA and ESEM. We first thoroughly evaluated the theoretical model (Simonds & Rothbart, 2004) as follows, using scoring criteria made available by Simonds and Rothbart (2009): (M1) higher order CFA with item level indicators

on latent first order factors, in turn loading onto latent second order factors to test the overall theorized structure; (M2) first order CFA to test the lower order structure only; (M3) A second order CFA using all a priori lower order scales to test the higher order structure only; (M4) first order ESEM with item indicators only to test for a complex structure in the proposed lower order structure; (M5) second order ESEM with all scale level indicators only complex structure in the proposed higher order structure.

We then moved to replication tests of lower order data-driven solutions beginning with Kotelnikova et al (2017), which we labeled supplemental due to their lack of theoretical basis. Here, we tested four models: (SM1) a CFA and (SM2) a geomin rotated ESEM of their 13 first order factors; (SM3) a CFA of their three second order factors; (SM4) a geomin rotated ESEM of their second order factors (2017). We then tested a higher order 3 factor solution from Nystrom and Bengtsson (2017) using a CFA that accepted their cross-loaded structure and specified correlated residuals (SM5). Finally, we evaluated results reported by Lipska et al. (2021). Here, we followed guidelines proposed by Golino et al. (2020) using the *EGAnet* package (v. 1.10; Golino & Christensen (2022) in R (v. 4.3.1; R-CoreTeam, 2022) verifying the fit of the model by re-expressing it as a CFA (SM6; Christensen, Gross, Golino, Silvia, & Kwapil, 2019). Results of these replication attempts were not intended to constrain our exploration of a theoretically aligned structure that would be applicable across typical and atypical populations.

**Phase 2: Exploration of a Theoretically Aligned Structure**—In this phase, a new item-level EFA initially to determine whether it could recover factors related to the theoretical articulation of the TMCQ. We conducted sensitivity analysis by varying the type of oblique rotations to ensure that any our proposed result would not be excessively method-dependent. For factor retention, we utilized parallel analysis that we compared to complex EFA models using the maximum likelihood in addition to considering model fit. We required a minimum standardized factor loading of 0.40 and allowed a maximum cross-loading of 0.30 for each model to consider a factor valid. For the retained factor model, we then evaluated each factor's internal consistency using both Cronbach's alpha and McDonald's omega with the *jmv* package (v. 2.0; Selker, Love, Dropmann, and Moreno (2021) in R (v. 4.3.1; R-CoreTeam, 2022). Once factors were deemed reasonably reliable, the revised (shortened) item list was subjected to CFA to challenge our hypothesis that cross loadings are necessary. We then moved to the primary ESEM model to address the expected model complexity. In a sensitivity analysis, the final lower order model was bootstrapped to evaluate likely generalizability. Factors recovered in the first order solution were entered into an ESEM model to fit a second order solution aimed at higher order factors as theorized.

**Phase 3: Measurement Invariance**—Models that were theoretically sound with adequate statistical fit were then subjected to measurement invariance testing across the ADHD and community control subsamples. For each possible solution, we began by fitting a multiple group model and then placing increasing invariance restrictions on the factor loadings and intercepts known as configural, metric, and scalar invariance (Chen, 2007; Putnick & Bornstein, 2016). In configural (weak) invariance, items are constrained to load on the same factors between groups. In metric invariance, items loadings are constrained to

be equivilent among the groups. In scalar (strong) invariance, item intercepts are constrained to be equivilent across groups. Models that continue to fit the data according to the above statistical fit criteria under increasing restriction and did not vary as restricts increase beyond previously established cutoffs ( CFI .01, RMSEA .015) were retained as invariant.

### Transparency and Openness

Raw data are publicly available at the NIH Data Archive (NDA). Analyses scripts and assembled data sets for re-analysis are available by emailing the corresponding author. These particular analyses were not preregistered.

## Results

### Phase 1: Tests of A Priori and Previously Proposed Factor Structures

Table 2 provides model fit information for all model tests. First, we tested the theorized higher order confirmatory model where the items load on their first order factors which in turn load onto three higher order latent variables—Effortful Control, Surgency, and Negative Affectivity (Simonds & Rothbart, 2009). This model was a poor fit to the data (M1 Table 2). We then tested a CFA of their 17 first order scale model, also with a poor fit (M2, Table 2). We used a CFA to test the three specified higher order factors represented by 14 lower order subscale means (Simonds & Rothbart, 2009), again with a poor fit (M3, Table 2). The ESEM test of their item level and scale level theoretical structures (M4, Table 2) yielded marginally acceptable fit, but many items did not load on their purported factors (as also observed by Kotelnikova et al., 2017). Further, some items loaded on their theorized target factors with equally high cross-loaded with non-target factors, making interpretation difficult. Table S1 provides the standardized factor loadings. Finally (M5, Table 2) an ESEM was fit that allowed subscale means to load freely on three second order factors, with unacceptable fit.

As a supplementary analysis, we next evaluated previously reported data-driven solutions beginning with Kotelnikova et al (2017). In the four models testing aspects of that result, only the ESEM replication of their first order solution was an acceptable fit (SM2, Table S2), although that solution was problematic for theoretical reasons. Table S2 provides fit statistics for all supplementary models. Table S3 contains the standardized factor loadings for the ESEM lower order solution of Kotelnikova et al., (2017). Table S3 shows that the items largely loaded on the same factors as they reported. However, the exceptions were notable and problematic. First, their factors had limited resemblance to the theorized structure. Second, their Fantasy factor was not recovered. Third, one lower order factor was characterized by two strong items cross-loadings with another factor (see Table S3). In all, a partial replication of Kotelnikova et al. (2017) was statistically supported but was deemed insufficient due to its poor match to theory. We then replicated Nystrom and Bengtsson's (2017) CFA model accepting the same cross-loadings and correlated residuals they had specified. The resulting model was a poor fit to our data (SM6, Table S2).

For our replication of Lipska et al. (2021), we represented the 17 lower order subscales (Simonds & Rothbart, 2009) in an EGA. Unlike them, however, we included Shyness in the model for theoretical reasons. Initial results suggested a four-factor structure, which we then

bootstrapped over 1000 iterations. Results suggested the four-factor solution was recovered 69% of iterations with the next most common result being a five-factor solution in 30% of iterations (see Figure S1). To obtain fit statistics for the four-factor model, we followed methodological guidance by re-expressing the modal result suggested by EGA as a CFA (Christensen et al., 2019). Fit for this model was poor (SM6, in Table S2, and Figure S2). See the supplementary material for further information regarding these analyses including the EGA dimension stability plot (Figure S3) and our discussion on the limitations of our replication attempts.

### Phase 2: Exploration of a Theoretically Aligned Factor Structure

**Item-Level Exploration using EFA**—To ensure results were not solely due to one rotation or estimation method, EFA models were rechecked across six models: geomin, oblimin, and promax rotations, each with both maximum likelihood robust and weighted least squares estimation methods. Solutions from the geomin and oblimin were similar, with the oblimin rotated model with robust maximum likelihood estimation method providing the simplest structure and the comparative best fit (M6, Table 2). Comparative fit statistics among all six of these EFA models are provided in the supplementary material (Table S4). Table S5 provides the factor loadings for the preferred oblimin robust maximum likelihood model. We proceeded with that model. Here, the parallel analysis suggested that eigenvalues up through the 14th factor were larger than the corresponding averaged eigenvalue for the randomized data sets, suggesting an initial 14 factor solution. Initial item trimming led us to discard one factor containing two only items (items 56 and 101; See Table S4), leaving a proposed 13 factors using 96 items. We labeled these 13 factors Activity, Affiliativeness, Anger, Assertiveness, Attentional Focus, Fear, Pain Sensitivity, Impulsivity, Reading, Openness, Perceptual Sensitivity, Sadness, and Shyness. Internal consistency statistics in the trimmed model were adequate (0.74 ≤ α ≤ 0.94) with the exception of the fear scale (α =.55). After efforts to salvage the fear scale with additional items, we made the decision to drop it, leaving 12 first order factors and 90 items[1]. Here, we also observed that one item (item 39; "can take a Band-Aid® off when needed, even when painful") negatively loaded on its target factor so we removed this item with minimal effect on scale reliability.

Table 3 contains omega reliability estimates for the originally articulated 17 scales (Simonds & Rothbart, 2004) and our retained 12 scales for the trimmed EFA model (M6, Table 2), which used 90 items. It is organized to show the alignment of the 12 retained scales with their theorized counterpart, and the final column of Table 3 shows the correlation of the retained scale with its originally proposed counterpart. That shows that the result to this point provides reasonable recovery of the content of many of those scales. The corresponding estimates for Cronbach's alpha are contained in the supplementary material in Table S6. A CFA of this 12-factor solution had inadequate fit (M7, Table 2), consistent with

---

[1]We attempted to salvage "fear" by reintroducing two items to the fear scale that had been trimmed previously (items 3 and 121) due to relatively low factor loadings, which improved internal consistency in Cronbach's alpha in the full sample (α =.73) and across subsamples (ADHD, α = .74; Non-ADHD, α = .72). However, inspection of McDonald's omega for the community subsample demonstrated less than adequate reliability (ω = .68). Given the difference may be a statistical artifact in light of the low factor loadings of items reintroduced into the scale, we made the decision to drop the fear scale from further analysis.

our hypothesis favoring complex models. We proceeded to the ESEM to test that hypothesis further.

**ESEM of the 12 Factor Trimmed Solution.:** We then evaluated an ESEM of the items in the prior (12-factor, 90 item) oblimin EFA solution, specifying the same trimmed 12 factors as before. As hypothesized, this more complex model provided a satisfactory fit to the data (M8, Table 2). Inspection of the individual factors revealed a pattern of small ($\lambda < .30$) yet significant ($p < .05$) cross-loadings; however, the theorized factor loadings remaining the largest loadings on target factors and were in the expected direction. Further supporting this solution, factor score correlations between the revised first order CFA (M7, Table 2) without cross-loadings and the final 12-factor ESEM (M8, Table 2) demonstrated a near identity between them ($r = .96 - .99$) as shown in Table 4. This suggested that the few cross-loadings we allowed in the ESEM model had minimal impact on factor distinctiveness or content meaning while substantially improving model fit and rendering it adequate.

**Sensitivity Analysis: ESEM-within-CFA Framework.**—Following methodological guidance (Marsh et al. 2014; Morin et al., 2013), to test the robustness of this last and best solution, we specified an ESEM-within-CFA (EwC) model to obtain bootstrapped parameter estimates (M8a, Table 2). This procedure involves re-expressing an ESEM solution as a CFA using the starting values generated from the ESEM model and constraining some indicators as "anchor items" while allowing other indicators to be estimated freely (Morin & Asparouhov, 2018). For anchor items, we selected items with large unstandardized loadings on their target factors with minimal cross-loadings on non-target factors and fixed these loadings on both target and non-target factors in the re-specified model. Residual variances were fixed to one. The model converged in 1000 re-samples and the baseline model was a good fit to our data thus providing some assurance that the final model is unlikely to be capitalizing on chance or local solutions excessively. The resulting unstandardized factor loadings and confidence intervals produced from the EwC analysis are provided in Table 5. We note that factor loadings for a small minority of items (< 10) on target factors fell below our initial trimming threshold in this model in the bootstrapped model; we nonetheless chose to retain these few items given substantive impact to reliability their removal would have had (see Table 5).

**Exploring the Higher Order Structure Based on our Lower Order Solution**—We tested an initial theoretical higher order model using ESEM where we selected the 12 scales theorized to load on the four second order factors of Surgency, Negative Affectivity, Affiliation, and Effortful Control as would be theoretically expected (Simonds & Rothbart, 2009). This model was a poor fit to the data (M9, Table 2). We then moved to a model with three higher order factors. To do so, we omitted the scales theoretically loading on Affiliation (Affiliation, Openness, and Assertiveness), and focused on scales theorized to load on Surgency, Negative Affectivity, and Effortful Control (Simonds & Rothbart, 2009). Allowing the nine scales to load freely on three latent variables improved model fit to acceptable standards (M10, Table 2) but with difficult-to-interpret factors, as follows. The first factor was characterized by significant loadings from Attentional Focus ($\lambda = 0.76$), Impulsivity ($\lambda = -0.71$), and again a factor that was a remnant of low intensity

pleasure, which we labeled Reading due to its item content ($\lambda = 0.55$), all suggesting an Effortful Control factor. That interpretation was somewhat clouded by a weak loading from Perceptual Sensitivity ($\lambda = 0.25$). Negative cross loadings from Sadness ($\lambda = -0.15$), and Anger ($\lambda = -0.23$) were also seen although they would not, in theory, be surprising in relation to Effortful Control factor. The second factor was characterized by primary factor loadings from Anger ($\lambda = 0.63$), Sadness ($\lambda = 0.69$), Pain Sensitivity ($\lambda = 0.52$), and Reversed Shyness ($\lambda = -0.60$), all suggesting a Negative Affectivity factor, although Reversed Shyness would have been expected to load on Surgency. It also had small negative cross-loadings from Attentional Focus ($\lambda = -0.15$), Activity ($\lambda = -0.29$), and Perceptual Sensitivity ($\lambda = 0.12$); although complicating the picture somewhat, these cross loadings are not theoretically inconsistent with an NA factor. The third factor was characterized by weak loadings in general, with the strongest from Perceptual Sensitivity ($\lambda = 0.45$), Impulsivity ($\lambda = 0.38$), Activity ($\lambda = 0.35$), Reading (again, a remnant of low intensity pleasure; $\lambda = 0.31$), Reversed Shyness ($\lambda = 0.28$), Anger ($\lambda = 0.27$), and Pain Sensitivity ($\lambda = 0.24$). The loadings from Impulsivity, Activity, Reversed Shyness, and Anger are all consistent with a weak Surgency factor, but a primary loading from Perceptual Sensitivity, low loadings overall, and multiple other items weaken the interpretation considerably.

The first two factors displayed adequate internal scale reliability ($\omega_1 = 0.70$, $\omega_2 = 0.75$) while reliability for the third factor was inadequate ($\omega_3 = 0.46$), consistent with its low factor loadings. Interpretively, it appeared to us the first two factors represented Effortful Control and Negatively Affectivity respectively though not without notable limitations, while these factors were dependent on a third, unreliable factor that only weakly related to Surgency. Overall, the higher order solution was perhaps reminiscent of the theoretical model but far from adequate in representing that model.

### Phase 3: Measurement Invariance

Finally, we tested for invariance between the ADHD and non-ADHD group in three models that had adequate fit to our data. These were the (a) 12 factor ESEM model we retained (M8), (b) the EwC model derived by our ESEM model (M8a), and (c) the higher order three factor ESEM model (M10), despite its questionable utility. The full results of invariance testing among the models are available in Table 6. All models met criteria for strong invariance between groups according to previously established cutoffs (Chen, 2007; Kotelnikova et al., 2017; Putnick & Bornstein, 2016). Additionally, we tested measurement invariance in our partial ESEM replication of Kotelnikova et al. (2017) where we were able to retain strong invariance over ADHD. Full results of this supplementary analysis are in Table S7.

## Discussion

The TMCQ is one of a series of instruments deriving from a leading model of child temperament, yet the theorized factor structure has not been reliably modeled, thus calling into question its construct validity and its place in child research. Satisfactory trait measurement is critical to advancing developmental theory of both temperament and personality, which are now accepted to be closely related and cover heavily overlapping

content domain at the questionnaire level (Shiner et al., 2021). Doing so is also important to understanding of the increasingly recognized trait-like aspects of developmental psychopathology. The present report is the largest sample to date to evaluate the TMCQ factor structure and expands on prior efforts by looking for the first time at factor structure both in a typical child sample and in a community-recruited sample of children with ADHD that was well-characterized by research evaluation procedures, many of whom had comorbid conditions. Our results support a 12-factor first order model that recovers, albeit with fewer items, the basic content of most of the original TMCQ scales. At the same time, important differences are notable, including the isolation of reading items from low intensity pleasure and the failure to reproduce key scales related to a putative Surgency higher order factor, notably High Intensity Pleasure. Perhaps because of these difference in lower-order factor structure, the theorized three-factor higher-order model (Effortful Control, Negative Affectivity, and Surgency) was only weakly reflected, with an unreliable third factor in the hypothesized Surgency space. Overall, results confirm the utility of a reduced set of items and restructured first-order scales in the TMCQ, but are problematic for the higher-order factor structure.

Methodologically, our results together with previous psychometric studies of the TMCQ (Kotelnikova et al., 2017; Lipska et al., 2021; Nystrom & Bengtsson, 2017) indicate that the TMCQ represents a solid yet incomplete foundation for working out the appropriate rating scale measurement of temperament in middle childhood. The present lower order result is more theoretically satisfying than prior alternatives solutions. Yet, the theorized higher order structure (Simonds & Rothbart, 2004) remains unable to be satisfactorily replicated across now four studies employing different measurement models. Indeed, it appears that satisfactory modeling of the TMCQ higher order structure requires modifications to the items to recapture needed contents and subscales, or modifying the theoretical structure. The more robust first order model depended on a small number of reasonable cross loadings. Such a complex structure in itself does not present serious problems. Historic conceptualizations of simple structure do not require that factors be completely orthogonal nor do they necessarily restrict all cross-loadings to be absolutely zero (Marsh et al., 2014). Additionally, the cross loadings retained here did not alter scale interpretation. The revised first order scale structure held within and across ADHD and non-ADHD samples, providing support for use in studies of both typical and at least some forms of atypical development. Even so, refinement of the item content appears to be in order, as the results yielded only 12 factors of which the remnant of low-intensity pleasure contained only items related to liking to read, and only 90 items were needed. This was similar to prior findings that the best solutions required only a subset of the items (Kotelnikova et al., 2017).

Results are also similar to Clark et al. (2020) who studied the Childhood Behavior Questionnaire (CBQ) that is intended for children ages 3–7. In their supplementary analysis, Clark et al. (2020) found an ESEM three-factor second order solution of the original CBQ scales was able to be retained in some of their validation samples, but noted unexpected cross-loadings in the three-factor higher-order structure. Similarly, recent psychometric papers on the self-reported EATQ-R have also found little evidence for a completely distinct higher order factor structure (Latham et al., 2020; Lawson, Atherton, & Robins, 2021). Though results from other parent report questionnaires have supported the broader higher

order structure of child temperament and personality in varying degrees in early and middle childhood (Neppl et al., 2010; Tackett, Krueger, Iacono, & McGue, 2008), we note more recent research in the personality literature has also suggested the relationship between these two sets of traits becomes more nuanced in middle childhood as children further develop their self-regulatory and self-evaluative capabilities (Shriner, 2021). This may challenge establishing the higher order traits with the TMCQ given its original formulation relied heavily on the CBQ in a laudable effort to retain some degree of measurement continuity across development (Rothbart, Ahadi, & Evans, 2000).

Yet by our reading, Rothbart's (2011) own conceptual model allowed for further differentiation of temperament traits as children age. One example that stands out from the current report concerned the fear scale, which was unreliable in our community sample. In our study, the highest loading items for fear were for situationally specific fears rather than reflecting a general tendency to fearfulness or being easily frightened. The items that were retained in our EFA (M6, Table 2) essentially show fears in three separate areas: loud noises, going fast, and heights; the remaining items that are theorized to load on the scale loaded poorly. Comparatively, the fear scale retained by Kotelnikova et al. (2017) had three items dealing with fear of the dark, burglars, and nightmares in addition to two highly correlated items dealing with the fear of needles loading on a separate factor all, again, situationally specific fears. However, Rothbart (2011) conceptualized the fear trait on the CBQ (from which the fear items on the TMCQ are derived) as "amount of negative affect including unease, worry, or nervousness related to anticipated pain or distress and/or potentially threatening situations (p. 51)." In light of this implicit intent to capture a general fearfulness trait, capturing a fear trait in middle childhood may warrant additional items (e.g., "startles easily," "is often frightened"). Also note that Rothbart's (2011) definition also encompasses anxiety and worry as much as fear. Anxiety and fear do seem to be differentiated by adulthood and perhaps become further differentiated to some extent in middle childhood (LoBue, Kim, & Delgado, 2019). Thus, it is interesting to consider whether temperament trait structure in middle childhood should seek to differentiate anxiety proneness from fearfulness at the lower-order factor level.

In the broader context of temperament development then, our results suggest that scale refinement of the TMCQ is necessary to better capture the higher order structure of temperament, whether in relation to Rothbart's (2011) theoretical definitions of temperament traits or a refined higher order model. This is especially true of Surgency, which has not been reliably or distinctly retained in any psychometric study of the TMCQ. Surgency is intended to capture construct related to extraversion (Evans & Rothbart, 2007) and is characterized by a disposition to positive emotionality, high activity level, and rapid approach to potential rewards (Rothbart, 2011, pp. 52–53). These traits are likely still present in middle childhood even though research from the personality literature suggests that mean stability of extraversion decreases as children age into early adolescence (e.g., Soto, 2016). Items to capture Surgency in this developmental area might then focus on identifying positive social rewards and self-evaluations (following Shriner, 2021) as well as prompts for parents to identify tendency towards exciting, but not always reckless behavior (e.g., "likes to skateboard", "tries to make others laugh").

The strength of the present research is that it entails a large comprehensive psychometric examination of both the original higher and lower models of the TMCQ in a large clinical and non-clinical sample recruited from the community. We offer a theoretically and psychometrically sound alternative for first order factors to use in future research. The restructured first order factors provide an improvement on the original in that in this sample, they were more psychometrically valid as judged by their fit to the data and yet appeared to have adequate theoretical construct validity in regard to their reasonable alignment with the rich theoretical history of the TMCQ for those particular scales. A principal limitation of this study, and all prior studies of this nature, is lack of independent sample replication. However, that said, we did find partial replication of Kotelnikova et al. (2017) as well as partial support for a revised lower order structure as theorized. Additionally, our findings did survive a rigorous sample permutation test of robustness, providing some reassurance against local over-fitting. The wider-than-recommended age range here (7–13 years) proved only a minor limitation; sensitivity analyses indicated that final results held when the sample was restricted to ages 7–10 (Table S8). Nevertheless, a major implication of our results is that the TMCQ can be used for ongoing child development research with modified but still highly meaningful versions of the lower order scales. The item set here clearly measures a portion of the intended theoretical constructs, and with refinement can be extended in that direction. Further, the modified structure shows measurement invariance in at least one clinical group—children with ADHD (many with comorbid conditions)—so results also provide support for continued use of a revised TMCQ in developmental psychopathology at least in some atypical populations.

Critically, results also confirm that additional revisions or supplemental measures would be beneficial to sharpen the coverage of the temperament domains and in particular to adequately represent a refined higher order structure, of which Rothbart and colleagues theorized three to four factors. Taking these results together with Kotelnikova et al. (2017), it appears quite possible to construct a briefer TMCQ using fewer than 100 items that has satisfactory validity and reliability for much of the original theoretical model. Indeed, continued scale development and revision in child personality has been crucial (Shiner et al., 2021); it has seemingly not been mirrored in the refinement of temperament scales in recent years. Doing so must balance the theoretical reality of both continuity and change and differentiation with development, as well as measurement continuity and difference. These findings suggest such refinement work for rating scales of temperament in middle childhood is now due.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## Reference:

Ablow JC, & Measelle JR (1993). The Berkeley puppet interview: Administration and scoring manuals.

Affrunti NW, Geronimi EM, & Woodruff-Borden J (2014). Temperament, peer victimization, and nurturing parenting in child anxiety: a moderated mediation model. Child Psychiatry Hum Dev, 45(4), 483–492. doi:10.1007/s10578-013-0418-2 [PubMed: 24202548]

Affrunti NW, & Woodruff-Borden J (2015). The associations of executive function and temperament in a model of risk for childhood anxiety. Journal of Child and Family Studies, 24(3), 715–724. doi:10.1007/s10826-013-9881-4

Asparouhov T, & Muthén B (2009). Exploratory structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 16(3), 397–438. doi:10.1080/10705510903008204

Ato E, Fernández-Vilar MÁ, & Galián MD (2020). Relation Between Temperament and School Adjustment in Spanish Children: A Person-Centered Approach. Front Psychol, 11, 250. doi:10.3389/fpsyg.2020.00250 [PubMed: 32140131]

Baker SR, & Victor JB (2001, August). African American children's personality: A confirmatory factor analysis. Paper presented at the European Conference of Development Psychology, Uppsala, Sweden.

Capaldi DM, & Rothbart MK (1992). Development and Validation of an Early Adolescent Temperament Measure. The Journal of Early Adolescence, 12(2), 153–173. doi:10.1177/0272431692012002002

Chen FF (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504. doi:10.1080/10705510701301834

Christensen AP, Gross GM, Golino HF, Silvia PJ, & Kwapil TR (2019). Exploratory Graph Analysis of the Multidimensional Schizotypy Scale. Schizophr Res, 206, 43–51. doi:10.1016/j.schres.2018.12.018 [PubMed: 30573405]

Clark DA, Donnellan MB, Durbin CE, Brooker RJ, Neppl TK, Gunnar M, . . . Putnam SP (2020). Using item response theory to evaluate the Children's Behavior Questionnaire: Considerations of general functioning and assessment length. Psychol Assess, 32(10), 928–942. doi:10.1037/pas0000883 [PubMed: 32584073]

Evans DE, & Rothbart MK (2007). Developing a model for adult temperament. Journal of Research in Personality, 41(4), 868–888.

Golino H, & Christensen AP (2022). EGAnet: Exploratory Graph Analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 1.0.1.

Golino H, Shi D, Christensen AP, Garrido LE, Nieto MD, Sadana R, . . . Martinez-Molina A (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. Psychol Methods, 25(3), 292–320. doi:10.1037/met0000255 [PubMed: 32191105]

Hattori M, Zhang G, & Preacher KJ (2017). Multiple Local Solutions and Geomin Rotation. Multivariate Behav Res, 52(6), 720–731. doi:10.1080/00273171.2017.1361312 [PubMed: 28952786]

Hu LT, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. doi:10.1080/10705519909540118

Inuggi A, Sanz-Arigita E, González-Salinas C, Valero-García AV, García-Santos JM, & Fuentes LJ (2014). Brain functional connectivity changes in children that differ in impulsivity temperamental trait. Front Behav Neurosci, 8, 156. doi:10.3389/fnbeh.2014.00156 [PubMed: 24834038]

Karalunas SL, Fair D, Musser ED, Aykes K, Iyer SP, & Nigg JT (2014). Subtyping attention-deficit/hyperactivity disorder using temperament dimensions: toward biologically based nosologic criteria. JAMA Psychiatry, 71(9), 1015–1024. doi:10.1001/jamapsychiatry.2014.763 [PubMed: 25006969]

Karalunas SL, Gustafsson HC, Fair D, Musser ED, & Nigg JT (2019). Do we need an irritable subtype of ADHD? Replication and extension of a promising temperament profile approach to ADHD subtyping. Psychol Assess, 31(2), 236–247. doi:10.1037/pas0000664 [PubMed: 30359050]

Kotelnikova Y, Mackrell SV, Jordan PL, & Hayden EP (2015). Longitudinal Associations Between Reactive and Regulatory Temperament Traits and Depressive Symptoms in Middle Childhood. J Clin Child Adolesc Psychol, 44(5), 775–786. doi:10.1080/15374416.2014.893517 [PubMed: 24742284]

Kotelnikova Y, Olino TM, Klein DN, Mackrell SVM, & Hayden EP (2017). Higher and Lower Order Factor Analyses of the Temperament in Middle Childhood Questionnaire. Assessment, 24(8), 1050–1061. doi:10.1177/1073191116639376 [PubMed: 27002124]

Latham MD, Dudgeon P, Yap MBH, Simmons JG, Byrne ML, Schwartz OS, . . . Allen NB (2020). Factor Structure of the Early Adolescent Temperament Questionnaire-Revised. Assessment, 27(7), 1547–1561. doi:10.1177/1073191119831789 [PubMed: 30788984]

Lawson KM, Atherton OE, & Robins RW (2021). The structure of adolescent temperament and associations with psychological functioning: A replication and extension of Snyder et al. (2015). J Pers Soc Psychol, 121(5), e19–e39. doi:10.1037/pspp0000380 [PubMed: 33539154]

Lipska A, Rogoza R, D bska E, Ponikiewska K, Putnam S, & Cieciuch J (2021). The structure of child temperament as measured by the Polish versions of the Children's Behavior Questionnaire and the Temperament in Middle Childhood Questionnaire: insight from the network psychometrics approach. Current Issues in Personality Psychology, 9(1).

LoBue V, Kim E, & Delgado M (2019). Fear in development. In Handbook of Emotional Development (pp. 257–282): Springer, Cham.

Marsh HW, Lüdtke O, Muthén B, Asparouhov T, Morin AJ, Trautwein U, & Nagengast B (2010). A new look at the big five factor structure through exploratory structural equation modeling. Psychol Assess, 22(3), 471–491. doi:10.1037/a0019227 [PubMed: 20822261]

Marsh HW, Morin AJ, Parker PD, & Kaur G (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. Annu Rev Clin Psychol, 10, 85–110. doi:10.1146/annurev-clinpsy-032813-153700 [PubMed: 24313568]

Marsh HW, Nagengast B, & Morin AJS (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. Dev Psychol, 49(6), 1194–1218. doi:10.1037/a0026913 [PubMed: 22250996]

Mervielde I, & De Pauw SSW (2012). Models of child temperament. In Zentner M & Shiner RL (Eds.), Handbook of temperament (pp. 21–40): The Guilford Press.

Morin AJS, & Asparouhov T (2018). Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA. Montreal, QC: Substantive Methodological Synergy Research Laboratory.

Morin AJS, Marsh HW, & Nagengast B (2013). Exploratory structural equation modeling. In Hancock GR & Mueller RO (Eds.), Structural equation modeling: A second course (pp. 395–436): IAP Information Age Publishing.

Muthén LK, & Muthén BO (1998–2017). Mplus User's Guide. (8th ed.). Los Angeles, CA: Muthén & Muthén.

Neppl TK, Donnellan MB, Scaramella LV, Widaman KF, Spilman SK, Ontai LL, & Conger RD (2010). Differential stability of temperament and personality from toddlerhood to middle childhood. J Res Pers, 44(3), 386–396. doi:10.1016/j.jrp.2010.04.004 [PubMed: 20634996]

Ng V, Cao M, Marsh HW, Tay L, & Seligman MEP (2017). The factor structure of the Values in Action Inventory of Strengths (VIA-IS): An item-level exploratory structural equation modeling (ESEM) bifactor analysis. Psychol Assess, 29(8), 1053–1058. doi:10.1037/pas0000396 [PubMed: 27736126]

Nigg JT, Karalunas SL, Gustafsson HC, Bhatt P, Ryabinin P, Mooney MA, . . . Wilmot B (2020). Evaluating chronic emotional dysregulation and irritability in relation to ADHD and depression genetic risk in children with ADHD. J Child Psychol Psychiatry, 61(2), 205–214. doi:10.1111/jcpp.13132 [PubMed: 31605387]

Nystrom B, & Bengtsson H (2017). A psychometric evaluation of the Temperament in Middle Childhood Questionnaire (TMCQ) in a Swedish sample. Scand J Psychol, 58(6), 477–484. doi:10.1111/sjop.12393 [PubMed: 28983922]

Posner MI, & Rothbart MK (2018). Temperament and brain networks of attention. Philos Trans R Soc Lond B Biol Sci, 373(1744). doi:10.1098/rstb.2017.0254

Putnick DL, & Bornstein MH (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. Dev Rev, 41, 71–90. doi:10.1016/j.dr.2016.06.004 [PubMed: 27942093]

Rothbart MK (1981). Measurement of temperament in infancy. Child Development, 52(2), 569–578. doi:10.2307/1129176

Rothbart MK (2011). Becoming who we are: temperament and personality in development. New York: Guilford Press.

Rothbart MK, Ahadi SA, & Evans DE (2000). Temperament and personality: origins and outcomes. J Pers Soc Psychol, 78(1), 122–135. doi:10.1037//0022-3514.78.1.122 [PubMed: 10653510]

Rothbart MK, Ahadi SA, Hershey KL, & Fisher P (2001). Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. Child Dev, 72(5), 1394–1408. doi:10.1111/1467-8624.00355 [PubMed: 11699677]

Rutter TM, & Arnett AB (2020). Temperament Traits Mark Liability for Coexisting Psychiatric Symptoms in Children With Elevated ADHD Symptoms. J Atten Disord, 1087054720943282. doi:10.1177/1087054720943282

Selker R, Love J, Dropmann D, & Moreno V (2021). jmv: The 'jamovi' Analyses: A suite of common statistical methods such as descriptives, t-tests, ANOVAs, regression, correlation matrices, proportion tests, contingency tables, and factor analysis. Version = 2.0. Retrieved from https://cran.r-project.org/web/packages/jmv/index.html

Shiner RL (2021). Personality development in middle childhood. In John OP & Robins RW (Eds.), Handbook of personality: Theory and research (pp. 284–302): The Guilford Press.

Shiner RL, Soto CJ, & De Fruyt F (2021). Personality Assessment of Children and Adolescents. Annual Review of Developmental Psychology, 3(1), 113–137. doi:10.1146/annurev-devpsych-050620-114343

Simonds J (2006). The Role of Reward Sensitivity and Response Execution in Childhood Extraversion. Retrieved from https://research.bowdoin.edu/rothbart-temperament-questionnaires/files/2016/09/Simonds_Dissertation_2006.pdf

Simonds J, & Rothbart MK (2004). The Temperament in Middle Childhood Questionnaire (TMCQ): A computerized self-report measure of temperament for ages 7–10. Paper presented at the Occasional Temperament, Athens, GA.

Simonds J, & Rothbart MK (2009). Scoring procedure, Temperament in Middle Childhood Questionnaire (TMCQ), version 3.0. Bowdoin College.

Soto CJ (2016). The Little Six Personality Dimensions From Early Childhood to Early Adulthood: Mean-Level Age and Gender Differences in Parents' Reports. J Pers, 84(4), 409–422. doi:10.1111/jopy.12168 [PubMed: 25728032]

Soto CJ, & John OP (2014). Traits in transition: the structure of parent-reported personality traits from early childhood to early adulthood. J Pers, 82(3), 182–199. doi:10.1111/jopy.12044 [PubMed: 23734942]

Tackett JL, Krueger RF, Iacono WG, & McGue M (2008). Personality in Middle Childhood: A Hierarchical Structure and Longitudinal Connections With Personality in Late Adolescence. J Res Pers, 42(6), 1456–1462. doi:10.1016/j.jrp.2008.06.005 [PubMed: 19122851]

Victor JB, Rothbart MK, & Baker SR (2003). Manual for the Child Temperament and Personality Questionnaire (CTPQ).

**Public Impact Statement:**

This study provides evidence for reliability and validity of a modified version the Temperament in Middle Childhood Questionnaire, a commonly used instrument in developmental and psychopathology research.

**Table 1**

Sample Characteristics

| Variable | Full Sample | ADHD | Community |
|---|---|---|---|
| *N* | 1,418 | 721 | 697 |
| *M* Age (*SD*) | 9.4(1.6) | 9.4(1.5) | 9.3(1.6) |
| Male % | 61% | 70% | 51% |
| Non-Hispanic white % | 78% | 79% | 78% |
| *M estimated* IQ (*SD*) | 109.2(14.1) | 106.7(14.2) | 111.8(13.5) |
| Median Income | $75,000–100,000 | $50,000–75,000 | $75,000–100,000 |
| Maternal report % | 82% | 84% | 80% |

Table 2

Fit Statistics for Model Tests of the TMCQ

| Model # | Description | χ² | BIC | CFI | TLI | RMSEA | RMSEA (90% CI) | SRMR | Statistical Evaluation |
|---|---|---|---|---|---|---|---|---|---|
| M1 | Higher order CFA w/ first and second order factors | 54161.87 | 574,710 | 0.62 | 0.61 | 0.05 | 0.049–0.050 | 0.18 | Poor |
| M2 | 17 first order factors CFA | 44059.21 | 564,544 | 0.71 | 0.70 | 0.044 | 0.043–0.044 | 0.21 | Poor |
| M3 | CFA w/ 3 second order factors and scale level indicators | 4369.76 | 37,547 | 0.59 | 0.50 | 0.202 | 0.197–0.207 | 0.18 | Poor |
| M4 | 17 first order factors ESEM Model | 18700.39 | 552,805 | 0.91 | 0.89 | 0.026 | 0.025–0.026 | 0.01 | Acceptable |
| M5 | 14 scale level ESEM w/ 3 second order factors | 1101.6 | 34,220 | 0.89 | 0.82 | 0.119 | 0.113–0.125 | 0.04 | Poor |
| M6 | Item level EFA with 14 retained first order factors | 24076.53 | 559,993 | 0.87 | 0.84 | 0.031 | 0.030–0.031 | 0.02 | Marginal |
| M7 | 12 first order factors CFA post item trimming | 13230.85 | 308,671 | 0.86 | 0.85 | 0.041 | 0.041–0.042 | 0.17 | Poor |
| M8 | 12 first order factors ESEM post item trimming | 6269.413 | 307,271 | 0.95 | 0.93 | 0.028 | 0.027–0.029 | 0.01 | Acceptable |
| M8a | 12 first order factors ESEM-within-CFA framework | 6815.96 | 307,752 | 0.94 | 0.92 | 0.03 | 0.029–0.031 | 0.04 | Acceptable |
| M9 | Rev. 12 Scale level ESEM w/ 4 SO Factors | 395.794 | 36,323 | 0.91 | 0.76 | 0.10 | 0.096–0.114 | 0.03 | Marginal |
| M10 | Rev. 9 Scale level ESEM w/ 3 SO Factors | 134.73 | 28,521 | 0.96 | 0.88 | 0.085 | 0.072–0.098 | 0.02 | Acceptable |

*Note:* All χ2 models are statistically significant at p < .001. FO = First Order, SO = Second Order. All ESEM model are rotated using a geomin rotation with 100 random starts.

**Table 3**

McDonald's Omega for Original and Revised Scales

| | Original TMCQ (17 Factors) | | | | | Revised TMCQ (12 Factors) | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Factor Label | N Items | Full Sample (N = 1,418) | ADHD (N = 721) | Community (N = 697) | Factor Label | N Items | Full Sample (N = 1,418) | ADHD (N = 721) | Community (N = 697) | r |
| Activ. Cont. | 15 | 0.81 | 0.77 | 0.77 | -- | -- | -- | -- | -- | -- |
| Activity | 9 | 0.90 | 0.91 | 0.89 | Activity | 10 | 0.91 | 0.87 | 0.90 | 0.97 |
| Affiliativeness | 10 | 0.83 | 0.8 | 0.83 | Affiliativeness | 9 | 0.83 | 0.81 | 0.83 | 0.96 |
| Anger | 7 | 0.88 | 0.87 | 0.85 | Anger | 9 | 0.88 | 0.88 | 0.87 | 0.93 |
| Assertiveness | 8 | 0.77 | 0.79 | 0.72 | Assertiveness | 4 | 0.76 | 0.78 | 0.69 | 0.88 |
| Attention | 7 | 0.95 | 0.87 | 0.92 | Attention | 11 | 0.94 | 0.87 | 0.92 | 0.99 |
| Discomfort | 10 | 0.81 | 0.81 | 0.80 | Pain sensitivity | 6 | 0.82 | 0.82 | 0.82 | 0.87 |
| Fear | 9 | 0.78 | 0.81 | 0.79 | -- | -- | | -- | -- | -- |
| HIP | 11 | 0.81 | 0.78 | 0.76 | -- | -- | | -- | -- | -- |
| Impulsivity | 13 | 0.93 | 0.82 | 0.77 | Impulsivity | 12 | 0.94 | 0.90 | 0.92 | 0.98 |
| Inhibit. Control | 8 | 0.84 | 0.90 | 0.91 | -- | -- | | -- | -- | -- |
| LIP | 8 | 0.67 | 0.72 | 0.79 | Reading | 3 | 0.82 | 0.82 | 0.79 | 0.76 |
| Openness | 9 | 0.80 | 0.79 | 0.82 | Openness | 6 | 0.79 | 0.78 | 0.79 | 0.92 |
| Perc. Sens. | 10 | 0.81 | 0.62 | 0.68 | Perc. Sens. | 10 | 0.83 | 0.83 | 0.84 | 0.95 |
| Sadness | 10 | 0.87 | 0.79 | 0.81 | Sadness | 4 | 0.87 | 0.88 | 0.87 | 0.86 |
| Shyness | 5 | 0.87 | 0.82 | 0.80 | Shyness | 6 | 0.87 | 0.87 | 0.85 | 0.97 |
| Soothability | 8 | 0.83 | 0.81 | 0.83 | -- | -- | | -- | -- | -- |
| | 157 | | | | | 90 | | | | |

*Note.* Pearson's r represents the correlation between the original and restructured scales when applicable.

Activ. Cont. is Activation Control, HIP is High Intensity Pleasure, LIP is Low Intensity Pleasure, Perc. Sens. is Perceptual Sensitivity.

**Table 4**

Factor Score Correlations between the CFA and ESEM factors

| CFA Factors | | | | | | ESEM Factors | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1. Attention | **0.99** | −0.76 | −0.01 | −0.45 | −0.08 | −0.02 | −0.28 | 0.09 | −0.29 | −0.06 | 0.41 | −0.04 |
| 2. Impulsivity | −0.76 | **0.99** | 0.18 | 0.52 | −0.11 | 0.07 | 0.30 | −0.05 | 0.29 | 0.16 | −0.32 | 0.29 |
| 3. Activity | −0.02 | 0.21 | **0.99** | −0.02 | −0.18 | 0.10 | −0.23 | 0.31 | −0.23 | 0.13 | −0.07 | 0.18 |
| 4. Anger | −0.50 | 0.59 | −0.02 | **0.97** | 0.31 | 0.19 | 0.61 | −0.12 | 0.68 | 0.06 | −0.20 | 0.29 |
| 5. Shyness | −0.07 | −0.14 | −0.21 | 0.31 | **0.97** | 0.05 | 0.27 | −0.40 | 0.46 | −0.20 | −0.11 | −0.28 |
| 6. Perceptual Sensitivity | 0.00 | 0.12 | 0.14 | 0.18 | 0.08 | **0.96** | 0.24 | 0.45 | 0.08 | 0.45 | 0.22 | 0.18 |
| 7. Pain Sensitivity | −0.29 | 0.32 | −0.24 | 0.60 | 0.29 | 0.21 | **0.99** | 0.05 | 0.51 | 0.10 | −0.08 | 0.09 |
| 8. Affiliation | 0.13 | −0.06 | 0.33 | −0.20 | −0.33 | 0.33 | 0.00 | **0.97** | −0.32 | 0.47 | 0.21 | 0.14 |
| 9. Sadness | −0.32 | 0.31 | −0.22 | 0.64 | 0.42 | 0.09 | 0.51 | −0.23 | **0.98** | −0.08 | −0.14 | 0.05 |
| 10. Openness | −0.03 | 0.18 | 0.13 | 0.05 | −0.16 | 0.40 | 0.10 | 0.49 | −0.10 | **0.98** | 0.42 | 0.26 |
| 11. Reading | 0.41 | −0.30 | −0.08 | −0.15 | −0.09 | 0.19 | −0.05 | 0.21 | −0.10 | 0.46 | **0.99** | 0.11 |
| 12. Assertiveness | −0.06 | 0.34 | 0.20 | 0.30 | −0.28 | 0.16 | 0.09 | 0.15 | 0.04 | 0.22 | 0.08 | **0.98** |

*Note.* Correlations > 0.05 are statistically significant at $p < .05$. Correlations above and below the diagonal represent the correlation between the ESEM and CFA factors only.

**Table 5**

Bootstrapped Unstandardized Factor Loadings for the ESEM-within-CFA (EwC) Model by Target Factor

| Scale | Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Attention | TMCQ149R | 0.92 | | | | | | | | | | | | -- |
| Attention | TMCQ120R | 0.91 | | | | | | | | | | | | 0.84 – 0.98 |
| Attention | TMCQ82R | 0.91 | | | | | | | | | | | | 0.83 – 0.99 |
| Attention | TMCQ80R | 0.96 | | | | | | | | | | | | 0.83 – 1.05 |
| Attention | TMCQ17R | 0.92 | | | | | | | | | | | | 0.80 – 1.03 |
| Attention | TMCQ78R | 0.85 | | | | | | | | | | | | 0.74 – 0.96 |
| Attention | TMCQ89R | 0.66 | | | | | | | | | | | | 0.54 – 0.76 |
| Attention | TMCQ93R | 0.63 | | | | | | | | | | | | 0.51 – 0.74 |
| Attention | TMCQ7R | 0.53 | | | | | | | | | | | | 0.43 – 0.66 |
| Attention | TMCQ70R | 0.46 | | | | | | | | | | | | 0.32 – 0.58 |
| Attention | TMCQ20 | 0.72 | | | | | | | | | | | | 0.57 – 0.85 |
| Impulsivity | TMCQ14 | | 0.78 | | | | | | | | | | | 0.64 – 0.92 |
| Impulsivity | TMCQ16 | | 0.81 | | | | | | | | | | | -- |
| Impulsivity | TMCQ22 | | 0.69 | | | | | | | | | | | 0.57 – 0.80 |
| Impulsivity | TMCQ25 | | 0.73 | | | | | | | | | | | 0.64 – 0.82 |
| Impulsivity | TMCQ72 | | 0.77 | | | | | | | | | | | 0.65 – 0.89 |
| Impulsivity | TMCQ74 | | 0.58 | | | | | | | | | | | 0.43 – 0.72 |
| Impulsivity | TMCQ79 | | 0.66 | | | | | | | | | | | 0.58 – 0.83 |
| Impulsivity | TMCQ83 | | 0.63 | | | | | | | | | | | 0.59 – 0.89 |
| Impulsivity | TMCQ108 | | 0.73 | | | | | | | | | | | 0.68 – 0.97 |
| Impulsivity | TMCQ124 | | 0.6 | | | | | | | | | | | 0.51 – 0.83 |
| Impulsivity | TMCQ130 | | 0.78 | | | | | | | | | | | 0.70 – 0.98 |
| Impulsivity | TMCQ42R | | 0.70 | | | | | | | | | | | 0.56 – 0.83 |
| Activity | TMCQ2 | | | 0.68 | | | | | | | | | | 0.60 – 0.76 |
| Activity | TMCQ21 | | | 0.68 | | | | | | | | | | 0.57 – 0.78 |
| Activity | TMCQ23 | | | 0.75 | | | | | | | | | | 0.65 – 0.85 |
| Activity | TMCQ37 | | | 0.75 | | | | | | | | | | 0.71 – 0.87 |
| Activity | TMCQ43 | | | 0.80 | | | | | | | | | | -- |

| Scale | Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Activity | TMCQ46 | | | 0.66 | | | | | | | | | | 0.59 – 0.73 |
| Activity | TMCQ66 | | | 0.51 | | | | | | | | | | 0.44 – 0.58 |
| Activity | TMCQ102 | | | 0.64 | | | | | | | | | | 0.58 – 0.71 |
| Activity | TMCQ115 | | | 0.52 | | | | | | | | | | 0.43 – 0.60 |
| Activity | TMCQ127 | | | 0.75 | | | | | | | | | | 0.69 – 0.82 |
| Anger | TMCQ146 | | | | 0.46 | | | | | | | | | 0.33 – 0.60 |
| Anger | TMCQ105 | | | | 0.65 | | | | | | | | | 0.48 – 0.89 |
| Anger | TMCQ100 | | | | 0.46 | | | | | | | | | 0.32 – 0.64 |
| Anger | TMCQ19 | | | | 0.53 | | | | | | | | | 0.37 – 0.73 |
| Anger | TMCQ24 | | | | 0.45 | | | | | | | | | 0.32 – 0.59 |
| Anger | TMCQ61 | | | | 0.57 | | | | | | | | | 0.43 – 0.71 |
| Anger | TMCQ87 | | | | 0.68 | | | | | | | | | – |
| Anger | TMCQ94 | | | | 0.73 | | | | | | | | | 0.61 – 0.84 |
| Anger | TMCQ99 | | | | 0.66 | | | | | | | | | 0.51 – 0.80 |
| Shyness | TMCQ28 | | | | | 0.52 | | | | | | | | 0.42 – 0.62 |
| Shyness | TMCQ47 | | | | | 0.59 | | | | | | | | 0.49 – 0.68 |
| Shyness | TMCQ55 | | | | | 0.86 | | | | | | | | – |
| Shyness | TMCQ118 | | | | | 0.88 | | | | | | | | 0.81 – 0.95 |
| Shyness | TMCQ59R | | | | | 0.64 | | | | | | | | 0.54 – 0.72 |
| Shyness | TMCQ26R | | | | | 0.49 | | | | | | | | 0.40 – 0.58 |
| Pain Sense | TMCQ5 | | | | | | | 0.68 | | | | | | 0.58 – 0.78 |
| Pain Sense | TMCQ30 | | | | | | | 0.56 | | | | | | 0.42 – 0.70 |
| Pain Sense | TMCQ63 | | | | | | | 0.44 | | | | | | 0.31 – 0.56 |
| Pain Sense | TMCQ64 | | | | | | | 0.46 | | | | | | 0.32 – 0.60 |
| Pain Sense | TMCQ138 | | | | | | | 0.77 | | | | | | – |
| Pain Sense | TMCQ154 | | | | | | | 1.04 | | | | | | 0.92 – 1.17 |
| Perc. Sense | TMCQ36 | | | | | | 0.55 | | | | | | | 0.43 – 0.65 |
| Perc. Sense | TMCQ44 | | | | | | 0.49 | | | | | | | 0.39 – 0.60 |
| Perc. Sense | TMCQ57 | | | | | | 0.52 | | | | | | | 0.44 – 0.61 |
| Perc. Sense | TMCQ77 | | | | | | 0.40 | | | | | | | 0.28 – 0.53 |
| Perc. Sense | TMCQ95 | | | | | | 0.38 | | | | | | | 0.29 – 0.47 |

| Scale | Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Perc. Sense | TMCQ109 | | | | | | 0.75 | | | | | | | -- |
| Perc. Sense | TMCQ111 | | | | | | 0.61 | | | | | | | 0.52 – 0.70 |
| Perc. Sense | TMCQ114 | | | | | | 0.65 | | | | | | | 0.54 – 0.75 |
| Perc. Sense | TMCQ123 | | | | | | 0.47 | | | | | | | 0.36 – 0.58 |
| Perc. Sense | TMCQ150 | | | | | | 0.55 | | | | | | | 0.44 – 0.65 |
| Sadness | TMCQ144 | | | | | | | | | 0.75 | | | | -- |
| Sadness | TMCQ133 | | | | | | | | | 0.74 | | | | 0.66 – 0.82 |
| Sadness | TMCQ97 | | | | | | | | | 0.70 | | | | 0.62 – 0.74 |
| Sadness | TMCQ27 | | | | | | | | | 0.65 | | | | 0.57 – 0.74 |
| Reading | TMCQ54 | | | | | | | | | | | 0.56 | | 0.46 – 0.65 |
| Reading | TMCQ73 | | | | | | | | | | | 0.87 | | -- |
| Reading | TMCQ86 | | | | | | | | | | | 0.94 | | 0.83 – 1.09 |
| Affiliation | TMCQ18 | | | | | | | | 0.56 | | | | | 0.45 – 0.68 |
| Affiliation | TMCQ33 | | | | | | | | 0.47 | | | | | 0.37 – 0.58 |
| Affiliation | TMCQ51 | | | | | | | | 0.44 | | | | | 0.35 – 0.54 |
| Affiliation | TMCQ58 | | | | | | | | 0.46 | | | | | 0.37 – 0.56 |
| Affiliation | TMCQ106 | | | | | | | | 0.52 | | | | | 0.39 – 0.64 |
| Affiliation | TMCQ129 | | | | | | | | 0.47 | | | | | 0.36 – 0.58 |
| Affiliation | TMCQ134 | | | | | | | | 0.38 | | | | | 0.28 – 0.49 |
| Affiliation | TMCQ148 | | | | | | | | 0.53 | | | | | 0.44 – 0.61 |
| Affiliation | TMCQ156 | | | | | | | | 0.68 | | | | | -- |
| Openness | TMCQ12 | | | | | | | | | | 0.36 | | | 0.28 – 0.44 |
| Openness | TMCQ48 | | | | | | | | | | 0.78 | | | 0.64 – 0.91 |
| Openness | TMCQ71 | | | | | | | | | | 0.40 | | | 0.26 – 0.52 |
| Openness | TMCQ104 | | | | | | | | | | 0.80 | | | -- |
| Openness | TMCQ116 | | | | | | | | | | 0.72 | | | 0.59 – 0.89 |
| Openness | TMCQ151 | | | | | | | | | | 0.35 | | | 0.25 – 0.45 |
| Assertiveness | TMCQ98 | | | | | | | | | | | | 0.61 | 0.49 – 0.73 |
| Assertiveness | TMCQ122 | | | | | | | | | | | | 0.4 | 0.34 – 0.49 |
| Assertiveness | TMCQ131 | | | | | | | | | | | | 0.67 | 0.58 – 0.78 |
| Assertiveness | TMCQ155 | | | | | | | | | | | | 0.69 | -- |

Author Manuscript

Author Manuscript

Author Manuscript

*Note:* CIs for anchor items are not produced in this model. Pain Sensitivity abbreviated to Pain Sense. Perceptual Sensitivity abbreviated to Perc. Sense.

Author Manuscript

**Table 6**

Results of Invariance Testing of First and Second Order Models between ADHD and Community Participants

| Model | Model Test | $\chi^2$ | BIC | CFI | TLI | RMSEA | SRMR | $\chi^2_{DIFF}$ | CFI | RMSEA |
|---|---|---|---|---|---|---|---|---|---|---|
| Rev. 12 FO Factors ESEM | Group Equiv. | 11180.16* | 306,603 | 0.930 | 0.92 | 0.029 | 0.03 | -- | -- | -- |
| | Configural | 10103.42* | 312,569 | 0.932 | 0.90 | 0.031 | 0.02 | -- | -- | -- |
| | Metric | 11054.10* | 307,037 | 0.931 | 0.92 | 0.031 | 0.03 | 1037.95* | -0.001 | 0.00 |
| | Scalar | 11180.16* | 302,130 | 0.930 | 0.92 | 0.029 | 0.03 | 126.07* | -0.001 | -0.002 |
| 12 FO Factors EwC | Group Equiv. | 11782.14 | 307,006 | 0.921 | 0.91 | 0.031 | 0.04 | -- | -- | -- |
| | Configural | 10628.35 | 312,904 | 0.923 | 0.89 | 0.033 | 0.03 | -- | -- | -- |
| | Metric | 11430.31 | 307,293 | 0.925 | 0.91 | 0.030 | 0.04 | 909.30 | 0.002 | -0.003 |
| | Scalar | 11577.58 | 306,880 | 0.924 | 0.91 | 0.030 | 0.04 | 147.44* | -0.001 | 0.000 |
| 3 SO Factors ESEM | Group Equiv. | 202.38* | 27,476 | 0.939 | 0.90 | 0.067 | 0.04 | -- | -- | -- |
| | Configural | 137.38* | 27,561 | 0.955 | 0.86 | 0.082 | 0.02 | -- | -- | -- |
| | Metric | 177.21* | 27,489 | 0.947 | 0.90 | 0.067 | 0.04 | 72.27* | -0.008 | -0.015 |
| | Scalar | 202.38* | 27,476 | 0.939 | 0.90 | 0.067 | 0.04 | 25.18* | -0.008 | 0.000 |

*Note:* FO = First Order, SO = Second Order.

*
Chi-square test was statistically significant at $p < 0.05$.

Difference tests are displayed for configural *vs.* metric and metric vs. scalar respectively.