Research on Child and Adolescent Psychopathology (2023) 51:1439–1451
https://doi.org/10.1007/s10802-023-01061-0

The P3 ERP in Relation to General Versus Specific Psychopathology in Early Childhood

Alexis Hosch[1] · **Jordan L. Harris**[1] · **Benjamin Swanson**[1] · **Isaac T. Petersen**[1]

Accepted: 20 March 2023 / Published online: 5 June 2023

Abstract
There is considerable covariation between externalizing and internalizing problems across the lifespan. Partitioning general and specific psychopathology is crucial to identify (a) processes that confer specific risk for externalizing versus internalizing problems and (b) transdiagnostic processes that confer risk for the covariation between externalizing and internalizing problems. The oddball P3 event-related potential (ERP) component, thought to reflect attentional orienting, has been widely examined in relation to psychopathology. However, prior studies have not examined the P3—or other aspects of neural functioning—in relation to general versus specific psychopathology in children. The present study examined whether children's ($N = 124$, ages 3–7 years) P3 amplitudes were associated with general versus specific psychopathology. Children's electroencephalography data were recorded during an oddball task. Parents rated their children's externalizing and internalizing problems. Using bifactor models to partition variance in parents' ratings of children's psychopathology symptoms, we examined children's P3 amplitudes in relation to three latent factors: (1) the general factor of psychopathology —the covariation of externalizing and internalizing psychopathology, (2) unique externalizing problems—the variance in externalizing problems after controlling for the general factor, and (3) unique internalizing problems. Results indicated that smaller P3 amplitudes were associated with unique externalizing problems at ages 3–5, and with general psychopathology at ages 6–7. Findings suggest that smaller P3 amplitudes may be associated with externalizing problems from a very young age. Moreover, there may be a developmental shift in the functional significance of the P3 in relation to general and specific psychopathology in childhood.

Keywords General psychopathology · P3 ERP · Bifactor · Externalizing · Internalizing · Children

Introduction

Psychopathology is pervasive across the lifespan. Externalizing behavior problems, including aggression, impulsivity, and oppositionality, as well as internalizing behavior problems, including depression, anxiety, and obsessions and compulsions, are among the symptoms most commonly experienced by children and adults (Forbes et al., 2016). Among children, attention-deficit/hyperactivity disorder (ADHD), anxiety problems, behavior/conduct problems, and depression are the most commonly diagnosed mental disorders, affecting between 4.4% and 9.8% of children (Bitsko et al., 2022). Furthermore, rates of mental illness among youth have been steadily increasing (Twenge et al., 2019). Moreover, psychopathology in childhood predicts severe outcomes across the lifespan, including academic underachievement (Hinshaw et al., 1992; Pedersen et al., 2019), substance use (Liu et al., 2011; Petersen et al., 2015), and criminality (White et al., 1990). Thus, it is crucial to understand mechanisms in the development of psychopathology, particularly during early childhood when these behaviors may be most amenable to intervention (Colizzi et al., 2020).

General Factor of Psychopathology

Externalizing and internalizing symptoms tend to be highly comorbid. A review found that oppositional defiant disorder, an externalizing disorder, was present along

Alexis Hosch and Jordan L. Harris contributed equally to this work.

✉ Alexis Hosch
alexis-hosch@uiowa.edu

[1] Department of Psychological and Brain Sciences, University of Iowa, 340 Iowa Avenue, Iowa City, IA 52242, USA

with an internalizing disorder in ~ 25% of preschoolers (Boylan et al., 2007). In children, externalizing and internalizing problems have been found to be correlated between $r = 0.40$ and 0.60 (Achenbach & Rescorla, 2001). The strong correlation and co-occurrence of internalizing and externalizing psychopathology has motivated modeling approaches and theoretical tests to determine whether there may be higher-order factors that account for the strong covariation of specific psychopathology (Kotov et al., 2017; Lahey et al., 2012). This higher-order factor is often referred to as the general factor of psychopathology, or general factor, and it accounts for what is common among internalizing and externalizing problems (Caspi et al., 2014; Lahey et al., 2012). It is also sometimes called the "p factor" when it additionally includes a thought disorder dimension, including symptoms of mania and psychosis (Caspi et al., 2014).

A bifactor approach is the most commonly used approach for estimating the general factor (Lahey et al., 2021). A bifactor model consists of questionnaire or structured interview items as indicators in a factor analytic model, loading onto one specific psychopathology factor, as well as loading directly onto an orthogonal general factor (e.g.,Clark et al., 2021; Hankin et al., 2017; Waldman et al., 2016). An example of a bifactor model is depicted in Fig. 1. The proliferation of general psychopathology models is due in part to a growing adoption of hierarchical nosologies, such as the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). Although HiTOP and other conceptualizations of general psychopathology are largely based on adult samples, several studies have replicated the presence of the general factor in younger children (e.g., McElroy et al., 2018; Olino et al., 2014).

A recent systematic review by Lynch et al. (2021) determined that there were several risk factors associated with the general factor as well as, or unique to, specific internalizing and externalizing factors in young people aged 10 to 24 years old. Among the risk factors unique to general psychopathology were: psychological variables (e.g., high negative affectivity), social factors (e.g., maternal depression), and biological variables (e.g., genetic risk for ADHD and schizophrenia, and reduced gray matter volume). Among the risk factors associated with the general factor were executive functioning deficits (Lynch et al., 2021). Notably, however, no studies identified by the Lynch review examined whether aspects of neural functioning are associated with general versus specific psychopathology.

Behaviors are highly heterogeneous in their mechanisms. The same set of behaviors can reflect different underlying substrates and can appear across several disorders. Thus, relying on behavior ratings is not sufficient to make conclusions about mechanisms in the development of psychopathology (Insel, 2014). Therefore, it is important to consider neural substrates of psychopathology, consistent with the National Institute of Mental Health's Research Domain Criteria (RDoC) initiative (Insel et al., 2010). The RDoC initiative encourages researchers to identify biological mechanisms of domains of psychopathology to

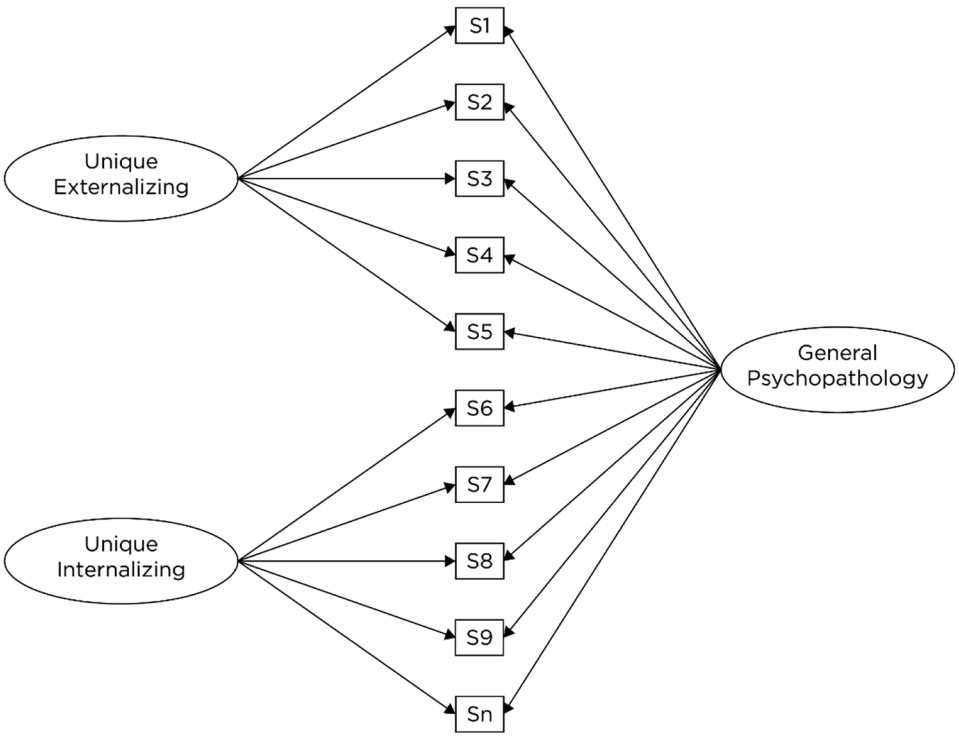


Fig. 1 Conceptual Depiction of a Bifactor Model of Psychopathology

map brain–behavior associations (Insel et al., 2010). Consistent with this approach, it may be important to consider executive-functioning-related neural processes in relation to general versus specific psychopathology. Executive functions refer to higher-order ("top-down") processes that exert control over attentional, cognitive, and behavioral tendencies in pursuit of a goal (Zhou et al., 2012). Executive dysfunctions, including difficulty with working memory or inhibitory control, are a key deficit in both externalizing and internalizing psychopathology (Schoemaker et al., 2013; Wang & Liu, 2021). One neural process that is thought to reflect attentional processes relevant for executive function is the oddball P3 ERP.

Oddball P3 ERP

The oddball P300 (or P3) event-related potential (ERP) has been widely studied as a neural process in psychopathology (Bernat et al., 2020; Bruder et al., 2011; Pasion et al., 2018). The P3 is often elicited in an oddball paradigm and is the third positive deflection of the ERP waveform. The P3 peaks between 400–600 ms post stimulus in young children, compared to around 300 ms in adults (Petersen et al., 2015; Polich, 2011). In adults, the P3 has been shown to have a central electrode distribution, whereas in children the distribution tends to be more parietal (Hoyniak et al., 2015; Polich, 2011). The P3 is thought to be generated by a neural circuit consisting of frontal, temporal, and parietal brain areas linked by dopaminergic neurotransmission (Pogarell et al., 2011; Polich, 2011). Task conditions, modality, and response type modulate the P3, but the P3 can also be elicited from a passive oddball paradigm in which participants do not make a behavioral response (Polich, 2011).

Multiple functional interpretations of the P3 exist. Context updating theory suggests that the P3 reflects neural activity associated with attentional orienting and working memory updating (Donchin et al., 1986; Polich, 2011). As memory and attention capabilities improve throughout childhood, peak auditory P3 amplitudes also increase, possibly reflecting the maturation of attentional resources over time (van Dinteren et al., 2014). Moreover, P3 latencies (i.e., the time interval between stimulus presentation and its peak) decrease across development, which may reflect more efficient cognitive processing (Gathercole, 1998; van Dinteren et al., 2014). Larger P3 amplitudes are thought to reflect increased neural activity in response to novel environmental stimuli (Polich, 2011), but the functional interpretation of the P3 may depend on the task. For example, in a passive oddball task (such as the one used in the present study), the P3 likely reflects attentional orienting to novelty (Friedman et al., 2001; Petersen et al., 2018). Attentional orienting—the ability to select and focus on specific information when multiple sensory stimuli are present in the environment—is a lower-level cognitive process necessary for carrying out higher-order (executive) functions (Posner & Petersen, 1990).

The P3 and Externalizing Problems

Deficient attentional orienting represents a key deficit in many individuals with externalizing problems (Dodge & Crick, 1990). According to social information processing theory, some people with externalizing problems are deficient in their ability to perceive whether others have hostile or aggressive intent (Crick & Dodge, 1994; Lansford et al., 2006). This failure to properly encode and subsequently process environmental stimuli may lead to an inappropriate situational response (Lansford et al., 2006). Given that the P3 is thought to reflect neural activity related to attentional orienting and working memory updating, it is possible that smaller P3 amplitudes may be a neurophysiological indicator of difficulties in the ability to orient to, and subsequently process, social environmental cues (Petersen et al., 2018). Indeed, meta-analytic work has shown that smaller P3 amplitudes elicited during oddball tasks are associated with externalizing psychopathology in children and adults (Gao & Raine, 2009). In adolescents, for example, smaller P3 amplitudes during a visual oddball task have been associated with externalizing psychopathology. Moreover, smaller P3 amplitudes in a passive auditory oddball task have been shown to predict within-person changes in aggression in young children (Petersen et al., 2018).

The P3 and Internalizing Problems

Although smaller P3 amplitudes have been widely associated with externalizing problems, findings on the association between the P3 and internalizing problems are less consistent. In adults, smaller P3 amplitudes may reflect cognitive slowing, consistent with the cognitive deficits seen in individuals with depression (Bruder et al., 2011). In support of this possibility, studies have shown that smaller P3 amplitudes are associated with depression in adults (for a review, see Bruder et al., 2011) and adolescents (Houston et al., 2003; Santopetro et al., 2022). By contrast, however, other studies have found no association between the P3 and depression in adolescents (Feldmann et al., 2018; Greimel et al., 2015), and another study found that *larger* P3 amplitudes were associated with depression in children (Lepistö et al., 2004).

Mixed findings have also emerged when examining the association between the P3 and anxiety. Theoretically, in the context of anxiety disorders, larger P3 amplitudes are thought to reflect an over-allocation of attention-related mechanisms (Reeb-Sutherland et al., 2009), particularly when perceiving threatening stimuli (Bechor et al., 2019). Indeed, larger

P3 amplitudes have been associated with anxiety in adolescence (Reeb-Sutherland et al., 2009). However, others have found that *smaller* P3 amplitudes are associated with anxiety (Bechor et al., 2019; Santopetro et al., 2022), or have found no association between the P3 and anxiety (Hogan et al., 2007). Given the inconsistent findings across studies of depression and anxiety, it is unclear how the P3 relates to internalizing psychopathology at a broadband level.

The P3 and General Psychopathology

To our knowledge, only one study has examined the association between the P3 ERP and general psychopathology, but this study was conducted in adults (Bernat et al., 2020). Bernat and colleagues (2020) found that smaller P3 amplitudes were related to both externalizing and internalizing problems, as well as the shared variance between externalizing and internalizing problems. The authors speculated that trauma experiences, or a shared genetic marker between externalizing and internalizing disorders, could explain this relation. When controlling for aggression, the association between smaller P3 amplitudes and general psychopathology was attenuated but remained. Thus, it is possible that smaller P3 amplitudes are associated with general psychopathology in children, but no studies have tested this possibility.

Gaps in Prior Research

In summary, very few studies have investigated neural processes in association with general and specific psychopathology. Bernat et al. (2020)—a study of adults—examined the P3 in relation to unique externalizing problems and general psychopathology, but not in relation to unique internalizing problems. Moreover, no study of neural processes has used bifactor modeling to partition general versus specific psychopathology, which is important for examining general and specific psychopathology in the same model to account for the covariation between externalizing and internalizing problems. Additionally, the relation between neural processes and general versus specific psychopathology has not been examined in children. The present study is the first to examine neural functioning in relation to general and specific psychopathology simultaneously in children. It is important to examine neural processes in relation to psychopathology in early childhood when neural processes supporting executive functions are rapidly developing.

The Present Study

The present study examined the association between the P3 ERP and general versus specific psychopathology in 3- to 7-year-olds. Consistent with prior research, we hypothesized that smaller P3 amplitudes would be associated with externalizing problems, in a bivariate association (Patrick et al., 2006; Petersen et al., 2018). Similarly, we hypothesized that smaller P3 amplitudes would also be associated with *unique* externalizing problems—i.e., specific psychopathology that remains after extracting the general factor. By contrast, given the inconsistent findings in children, we had no hypotheses regarding the association between P3 amplitudes and internalizing problems, in a bivariate association. We expected that any associations of the P3 with internalizing problems would be driven by the general factor. Thus, we hypothesized that P3 amplitudes would not be associated with unique internalizing problems. Consistent with findings in adults (Bernat et al., 2020), we hypothesized that smaller P3 amplitudes would be associated with general psychopathology, i.e., the general factor.

Method

Participants

Participants consisted of a community sample of young children ($N = 124$, $M_{age} = 4.80$ years, $SD = 1.20$ years, 59 girls) and their families, who took part in an ongoing accelerated longitudinal study. Children were recruited from 2018–2022 at one of the following ages: 36 ($n = 34$), 45 ($n = 34$), 54 ($n = 24$), or 63 ($n = 32$) months and were assessed every 9 months over 4 time points. The full sample of children spanned 3 to 7.5 years of age. Participants were recruited from Iowa City, Iowa and surrounding areas. Exclusion criteria were: the child's primary caregiver did not speak English, or the child did not have a permanent guardian, did not have normal or corrected-to-normal vision and hearing, or was not capable of following basic instructions in English. See Supplementary Fig. S1 for a detailed enrollment flowchart.

The final sample consisted of the children, their primary caregiver, and the primary caregiver's parenting partner (as applicable). The racial composition of children in the sample was: 75.0% White, 7.3% Black or African American, 5.6% Asian, 6.5% multiracial, 5.6% other. The ethnic composition of the sample was 12.9% Hispanic or Latino and 87.1% Non-Hispanic or Latino. Participant demographics are detailed in Supplementary Appendix S1. Compared to the U.S. population, participants in the sample were somewhat more likely to be Non-Hispanic White, married, be middle or upper class, and have a college or graduate degree. Participant demographics were broadly reflective of the surrounding area.

Procedure

At each time point (i.e., every 9 months for four time points), the child and their primary caregiver completed two lab

visits, approximately one week apart. Informed consent was obtained from the primary caregiver during the first lab visit. The primary caregiver completed electronic questionnaires during both lab visits or from home. The primary caregiver's parenting partner was emailed or mailed the questionnaires to complete. During the first lab visit, the child completed a series of behavioral tasks, which are not the focus of the present study. During the second lab visit ($M = 86.6$ min, $SD = 18.8$), the child completed computerized tasks, including an oddball task, while wearing an electroencephalography (EEG) cap and brainwaves were recorded.

Measures

The present study is part of a larger study, the School Readiness Study. Hypotheses and measures for the School Readiness Study were pre-registered: https://osf.io/jzxb8. Hypotheses methods, and a data analysis plan for the present study were co-registered: https://osf.io/pny26. Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/zs2bn. Descriptive statistics and a correlation matrix are in Table 1. Reasons for missingness and tests of systematic missingness are in Supplementary Appendix S2.

Behavior Problems

Achenbach System of Empirically Based Assessment (ASEBA)

The ASEBA assesses children's emotional and behavioral problems. Items were rated on a 3-point Likert scale according to how well the item described the child (0 = not true, 1 = somewhat or sometimes true, 2 = very true). Parents completed the Child Behavior Checklist (CBCL) 1.5–5 (Achenbach & Rescorla, 2000) if the child was 3–5 years old or the CBCL 6–18 (Achenbach & Rescorla, 2001) if the child was 6–7 years old. Scores from items on the Externalizing and Internalizing scales were used. The ASEBA scales are empirically derived, widely used, and have shown strong reliability (internal consistency, test–retest reliability, and interrater reliability) and validity (content, construct, and criterion-related validity) in large and diverse samples in the U.S. (Sattler, 2014). Reliability estimates of ratings are in Supplementary Appendix S3.

Oddball Task

We used a passive auditory oddball task (Petersen et al., 2018), which is often used with EEG to assess novelty detection and

Table 1 Correlations among predictors, outcomes, and covariates

Variable	Sex	Age	SES	Trials Kept	Bad Channels	P3	EXT	INT
Sex	—							
Age	0.09†	—						
SES	-0.01	0.13*	—					
Trials Kept	0.05	0.14*	0.06	—				
Bad Channels	0.04	-0.07	-0.07	0.24***	—			
P3	-0.22***	0.24***	0.18***	0.01	-0.09	—		
EXT	-0.15***	-0.29***	-0.15***	0.04	0.04	-0.16*	—	
INT	0.14*	0.00	-0.11*	0.01	0.05	-0.08	0.64***	—
M	0.48	4.81	-0.07	20.79	8.37	2.78	0.15	0.07
SD	0.50	1.21	0.86	5.79	18.32	3.69	0.14	0.06
Min	0.00	2.92	-3.47	13.00	1.00	-6.15	0.00	0.00
Max	1.00	7.80	3.23	47.00	128.00	18.25	0.77	0.40

"Age" in years. "Sex" is coded such that 1 = female and 0 = male. "SES" = socioeconomic status. "Trials Kept" is the final number of trials not excluded during the oddball P3 task. "Bad Channels" is the count of channels excluded from analysis due to poor quality. "P3" is the child's P3 amplitude on the infrequent trials of the oddball task. For the table, externalizing and internalizing problem scores were converted to a proportion of the maximum possible score to put scores onto a metric with the same possible range, such that higher scores reflected more problems. In the correlation matrix, "EXT" and "INT" represent composite scores of mothers' and fathers' reports of externalizing and internalizing problems, respectively, on the Child Behavior Checklist (CBCL) in long format (each participant has a row for each combination of participant, wave, and rater). Descriptive statistics for externalizing and internalizing problems are based on mothers' ratings on the CBCL. Descriptive statistics for fathers' ratings of externalizing problems were $M(SD) = 0.18(0.14)$, range = 0–0.62; and for internalizing problems were $M(SD) = 0.08(0.06)$, range = 0–0.33. "Min" = lowest score in the sample. "Max" = highest score in the sample

† $p < 0.10$; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.001$, all *p*s are two-tailed

attention. The task was administered on a computer using E-Prime software (version 2.0.10.356; Schneider et al., 2012). During the task, the child silently viewed an age-appropriate television show on an android tablet. Pure, low-frequency (1000 Hz) and high-frequency (1500 Hz) tones were presented from speakers on both sides of the computer monitor. The task consisted of 120 randomized trials, including 84 frequent stimuli (i.e., tones) and 36 infrequent stimuli. The tone that was the infrequent versus frequent stimulus was counterbalanced across participants, and infrequent tones did not repeat on consecutive trials. Each tone lasted approximately 300 ms, and the interstimulus interval varied randomly from 2300–2500 ms to prevent habituation.

Electrophysiological Recordings and Data Processing

Electrophysiological data were collected using an Electrical Geodesic, Inc (EGI) 128-electrode Hydrocel Geodesic Sensor Net with a Net Amps 400 series amplifier. Net Station Acquisition Software 5.4.2 (Electrical Geodesics, Inc., 2018) was used to collect the continuous EEG data. A detailed description of the collection and pre-processing of the EEG data is in Supplementary Appendix S4.

After pre-processing, ERP waveforms were statistically decomposed using sequential temporospatial principal components analysis (tsPCA), which empirically identifies the regions of electrodes and time frames that account for the most variance in the waveforms. The tsPCA was conducted using the ERP PCA Toolkit (version 2.97; Dien, 2010). We performed tsPCA separately for each condition (i.e., frequent versus infrequent trials), consistent with prior findings that a combined PCA (i.e., including both trial types in one PCA) misallocated substantial variance (Barry et al., 2018). For sensitivity analyses, we also performed tsPCA separately for two age groups across conditions, consistent with Scharf et al. (2022). Self-regulation and executive functioning abilities rapidly increase from ages 3 to 6 (Diamond, 2002; Hosch et al., 2022; Montroy et al., 2016), which may produce different PCA component structures in younger versus older children. To retain comparable sample sizes for obtaining robust PCA results, "younger" children were defined as 36–54 months ($n = 94$ observations) and "older" children were defined as 63–90 months ($n = 65$ observations). The tsPCA results, including discussion of tsPCA components, are in Supplementary Appendix S5.

The tsPCA component structure appeared different in younger versus older children. In addition, P3 amplitudes from the overall tsPCA (i.e., not separated by age) were weakly correlated with P3 amplitudes from the age-specific tsPCA ($r[155] = 0.44$). Thus, to account for neurodevelopmental changes, P3 amplitudes from the age-specific tsPCA were used in analyses. The grand average waveform for each age group is depicted in Fig. 2. The chosen temporospatial components peaked between 470–520 ms for younger children and 440–455 ms for older children. The temporospatial components were characterized by a posterior positivity that differentiated frequent and infrequent trials (in Supplementary Figs. S2 and S3). The P3 tsPCA components were extracted using the AutoPCA function. Participants' peak amplitudes of the P3 tsPCA component were used in analyses. The P3 tsPCA component waveforms, topo-plots, and electrode clusters are in Supplementary Figs. S4–S8. The child's P3 amplitude on infrequent trials was used as the predictor variable. Children's P3 amplitudes showed moderate 9-month cross-time stability ($r[53] = 0.31$, $p < 0.001$).

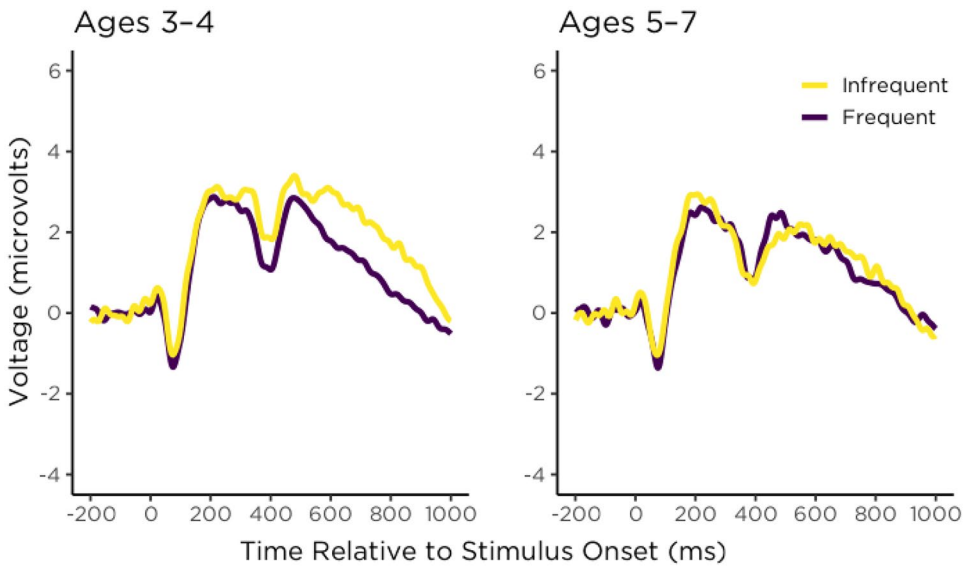


Fig. 2 Grand-Averaged Waveform

Covariates

We examined models with and without covariates. Covariates included: the child's age, child's sex (male = 0, female = 1), family socioeconomic status, informant type (mother = 0, father = 1), the number of infrequent trials kept, and the number of "bad" trials. A full description of covariates is in Supplementary Appendix S6.

Statistical Analysis

We fit a bifactor model to determine whether P3 amplitudes predicted general versus specific psychopathology. First, we fit a bifactor model with no predictors that included a latent factor for the general factor of psychopathology, in addition to latent factors for externalizing problems and internalizing problems. The latent factors were set to be uncorrelated, so the general factor represented the covariation among all externalizing and internalizing items. By contrast, the specific psychopathology factors—i.e., externalizing problems and internalizing problems—represented the covariation among the items within that dimension after extracting the variance accounted for by the general factor. That is, the externalizing latent factor represented unique externalizing variance, and the internalizing latent factor represented unique internalizing variance. Separate bifactor models were fit to ages 3–5 (CBCL 1.5–5) and ages 6–7 (CBCL 6–18), because the measures had different item content.

Bifactor models were fit in Mplus version 8.6 (Muthén & Muthén, 2021). Because the behavior problem items were ordinal, models used a robust diagonally weighted least square mean and variance adjusted estimator (WLSMV). The latent factors were standardized to have a mean of 0 and a standard deviation of 1. Items were allowed to load freely on the relevant factors. Missing data were handled with pairwise deletion, the default setting for WLSMV, which maximizes the use of available data. We also conducted sensitivity analyses using multiple imputation. We followed a stepwise process to fit a bifactor model, as described in Supplementary Appendix S7. Upon establishing a well-fitting bifactor model, we added children's P3 amplitudes as a predictor. Then, we added model covariates. The predictor and covariates were allowed to predict the three latent factors.

We examined model fit in root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and standardized root mean square residual (SRMR). Model fit was considered good if RMSEA ≤ 0.05, CFI ≥ 0.95, TLI ≥ 0.95, SRMR ≤ 0.08; model fit was considered acceptable if RMSEA ≤ 0.08, CFI ≥ 0.90, TLI ≥ 0.90, SRMR ≤ 0.10 (Bentler & Bonett, 1980; Hu & Bentler, 1999; Schermelleh-Engel et al., 2003; Schreiber et al., 2006). To determine the relative strength of the general factor, we calculated the explained common variance (ECV) in ratings of psychopathology (see Supplementary Appendix S8).

Sensitivity Analyses

We conducted several sensitivity analyses. To account for missingness in the data, we performed multiple imputation. To test whether results were specific to neural activity in response to infrequent stimuli, we also considered P3 amplitudes to frequent stimuli.

In addition, we conducted sensitivity analyses with P3 amplitudes computed as a difference score (infrequent minus frequent amplitudes). As a general note, we interpret the difference score models with caution, because difference scores tend to be less reliable when (a) the individual indices are correlated (Revelle & Condon, 2019), and when (b) the individual indices have similar variances (Trafimow, 2015), both of which were the case in the present study. Moreover, using difference scores examines a different conceptual question, i.e., the ability to distinguish between environmental stimuli, than examining neural activity in response to infrequent stimuli. For these reasons, we focus our interpretations on P3 amplitudes from infrequent trials in relation to general and specific psychopathology.

Results

Bivariate Associations

Children's P3 amplitudes were negatively associated with their parent-reported externalizing problems. That is, smaller P3 amplitudes were associated with externalizing problems. No significant association was found between children's P3 amplitudes and their parent-reported internalizing problems.

Bifactor Model

Deriving The Final Model

We started by fitting a bifactor measurement model. A bifactor model with all externalizing items (ages 3–5: 24 items; ages 6–7: 35 items) and internalizing items (ages 3–5: 36 items; ages 6–7: 32 items) did not converge. Thus, we followed a stepwise process to derive the final model, as described in Supplementary Appendix S9. The predictor (P3 amplitudes) and covariates were added to the final model. Factor loadings of the items are in Supplementary Tables S1–S2. ECV and ECVs estimates for both models are in Supplementary Fig. S9. Sensitivity analyses are in Supplementary Appendix S10.

Ages 3–5 Model

For children ages 3–5, model fit indices were strong according to RMSEA (0.027), acceptable according to CFI (0.931) and TLI (0.926), and were somewhat weaker according to SRMR (0.112). Regression coefficients of predictors and covariates are in Supplementary Table S3. P3 amplitudes were negatively associated with unique externalizing problems (β = -0.24, $p = 0.015$). However, P3 amplitudes were not significantly associated with unique internalizing problems (β = 0.11, $p = 0.348$), or the general factor (β = -0.06, $p = 0.454$). Findings held when using multiple imputation. Results are depicted in Fig. 3.

P3 amplitudes on frequent trials were not associated with general or specific psychopathology. P3 amplitude difference scores were negatively associated with the general factor (β = -0.17, $p = 0.034$), but were not associated with unique externalizing or unique internalizing problems.

After accounting for covariates, the association between P3 amplitudes and unique externalizing problems remained significant. We also found that unique externalizing problems decreased with age, whereas unique internalizing problems increased with age. Boys showed more unique externalizing problems than girls, whereas girls showed more unique internalizing problems than boys. The general factor was not significantly related to the child's age or sex.

Ages 6–7 Model

For children ages 6–7, model fit indices were acceptable according to RMSEA (0.041), CFI (0.929), and TLI (0.918), and were somewhat weaker according to SRMR (0.147). Regression coefficients of predictors are in Supplementary Table S4. P3 amplitudes were negatively associated with the general factor (β = -0.39, $p = 0.018$). However, P3 amplitudes were not significantly associated with unique externalizing problems (β = -0.16, $p = 0.384$) or unique internalizing problems (β = 0.47, $p = 0.189$). Results are depicted in Fig. 3. Larger P3 amplitudes on frequent trials were associated with greater unique internalizing problems (β = 0.37, $p = 0.049$), but not the general factor or externalizing problems.

We added covariates to the model to determine whether the association between P3 amplitudes and the general factor would remain. The model did not converge with all covariates. Thus, we examined each covariate in separate models to determine which ones were associated with the general factor. Only the child's sex (β = 0.43, $p < 0.001$) was significantly associated with the general factor, such that girls showed higher levels than boys. However, a model that included both P3 amplitudes and the child's sex did not converge. Thus, we examined the P3 in relation to general versus specific psychopathology controlling for each covariate separately. In each case, the model either did not converge

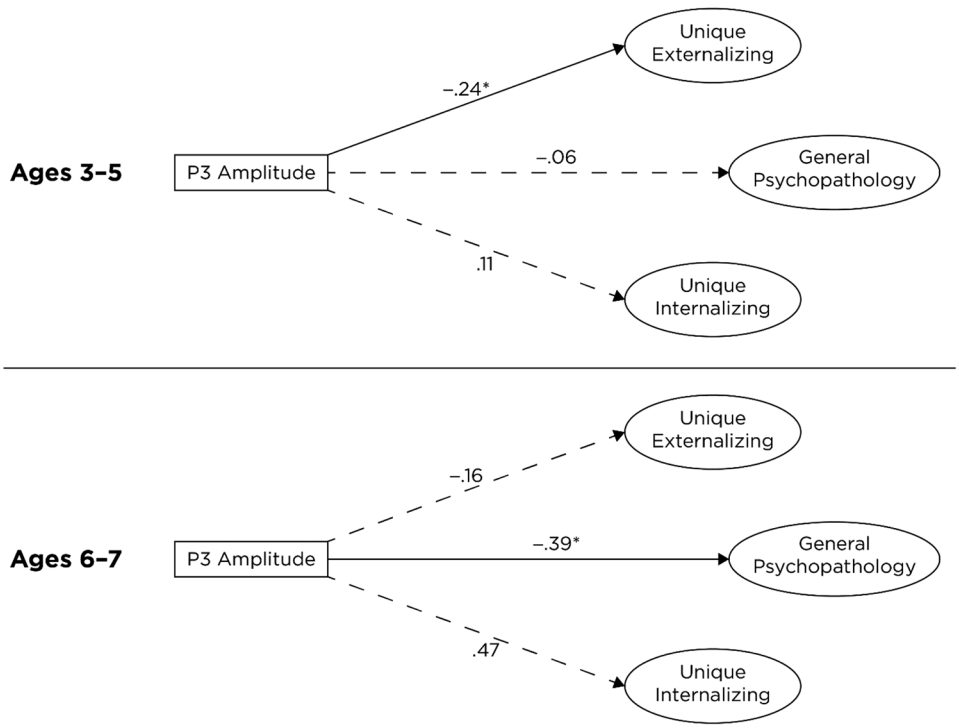


Fig. 3 Results of the Bifactor Model

or the P3 remained associated with the general factor. The general factor was not significantly related to the child's age.

Despite a non-significant association due to somewhat larger standard errors, the effect size of the association between P3 amplitudes and the general factor (β = -0.37) in the model with multiple imputation was similar to the effect size in the model without multiple imputation. P3 amplitude difference scores were not associated with the general factor, or with unique externalizing or unique internalizing problems.

Discussion

As expected, the present study found that smaller P3 amplitudes were associated with unique externalizing problems at ages 3–5 and the general factor of psychopathology at ages 6–7, with medium effect sizes. Consistent with hypotheses, P3 amplitudes were not significantly associated with unique internalizing problems in either age group. Contrary to hypotheses, however, smaller P3 amplitudes were not significantly associated with the general factor at ages 3–5 or unique externalizing problems at ages 6–7. These results suggest a developmental shift in the functional significance of the P3 amplitude in early childhood.

In the younger children (ages 3–5), smaller P3 amplitudes were associated with externalizing problems. Previous studies have identified smaller P3 amplitudes in association with externalizing problems among children as young as 2.5 years of age (Petersen et al., 2018). However, the present study is the first to demonstrate an association between P3 amplitudes and *unique* externalizing problems, controlling for the covariation with internalizing problems. In the older children (ages 6–7), smaller P3 amplitudes were associated with the general factor. This replicates previous findings in adults that smaller P3 amplitudes are associated with higher levels on a factor representing the shared variance between externalizing and internalizing problems (Bernat et al., 2020). The association between P3 amplitudes on infrequent trials and outcomes was not replicated when examining P3 amplitudes on frequent trials. Thus, the associations observed between smaller P3 amplitudes and psychopathology did not merely reflect generally attenuated neural processing.

We also observed associations of the latent factors with the child's age and sex. Unique externalizing problems decreased across ages 3–5, whereas unique internalizing problems increased across ages 3–5. At ages 3–5, boys showed more unique externalizing problems compared to girls, whereas girls showed more unique internalizing problems compared to boys. At ages 6–7, girls showed higher levels on the general factor than boys.

Interpretation of Findings

Results from the present study suggest that smaller P3 amplitudes may be an early neural marker of psychopathology in children. The P3 ERP component is thought to reflect neural processes related to attentional orienting to novelty (Donchin et al., 1986; Polich, 2011). Smaller P3 amplitudes may, therefore, reflect deficits in attentional orienting. In younger children (ages 3–5), deficits in attentional orienting (as indexed by smaller P3 amplitudes) may lead to externalizing behavior. By contrast, deficits in attentional orienting may lead to a broader range of psychopathology in older children (ages 6–7). These interpretations are consistent with prior evidence of associations between smaller P3 amplitudes and externalizing (Gao & Raine, 2009; Pasion et al., 2018; Patrick et al., 2006) and general psychopathology (Bernat et al., 2020).

Interestingly, we also observed that larger P3 amplitudes on frequent trials were associated with greater unique internalizing problems at ages 6–7. To our knowledge, this is the first study to observe this association. Speculatively, children with internalizing problems may be more likely to over-allocate attentional resources (as indexed by larger P3 amplitudes) to stimuli. Indeed, research has shown that some children with anxiety show attentional biases to threat (Valadez et al., 2022). Notably, the stimuli used in this study did not have an emotional valence. However, our findings suggest that children with internalizing problems (e.g., anxiety) may over-attend to stimuli more generally, even when non-threatening.

It is unclear, however, why smaller P3 amplitudes were associated with unique externalizing problems, but not general psychopathology, in younger children and with general psychopathology, but not unique externalizing, in older children. There appears to be a developmental shift that occurs between ages 3 and 7, in which smaller P3 amplitudes become more strongly associated with general psychopathology over time. Speculatively, there may be different consequences of deficient attentional processes depending on the developmental period. For instance, consequences of deficient attentional orienting may be more limited to disruptive behavior in early childhood. By contrast, deficient attentional processing may have consequences for a broader range of impairment in older children, consistent with what is observed in adults (Bernat et al., 2020).

However, it is also possible that attentional deficits are, in fact, related to general psychopathology in young children, but are only detected in relation to externalizing problems because of how internalizing problems manifest in early childhood. Younger children may be more likely than older children to express internalizing symptoms, e.g., sadness, as externalizing behaviors, e.g., aggression or emotional outbursts (Kopp, 1982). Research has shown

that self-regulation abilities rapidly increase from ages 3–6 then slow and level off from ages 6–7 (Diamond, 2002; Hosch et al., 2022; Montroy et al., 2016). As children develop and their self-regulatory skills improve, internalizing behaviors may appear more prototypical, as externalizing behaviors are less likely to be expressed. Consistent with this explanation, unique internalizing problems increased from ages 3–5.

In the older children (ages 6–7), the two items with the strongest loading on the general factor (among items that also loaded onto a specific factor) were "screams a lot" ($\beta = 0.95$) and "temper tantrums or hot temper" ($\beta = 0.77$). These behaviors are aspects of a broader behavioral construct, negative emotionality, which is highly correlated with general psychopathology (Forbes et al., 2016). Negative emotionality refers to the tendency to display negative emotions, such as anger or sadness, and is thought to capture behavioral aspects of both internalizing and externalizing problems. Our findings suggest that children with smaller P3 amplitudes may be more likely to display a broader range of behaviors reflective of negative emotionality, as opposed to specific externalizing behaviors, as they get older. This could explain the apparent developmental shift, in which P3 amplitudes appear to become more strongly associated with the general factor at later ages. Ultimately, more research is needed to clarify the developmental relation between P3 amplitudes and general versus specific psychopathology.

Strengths

The present study had key strengths. First, the study examined two age groups in early childhood using a community sample. Second, multiple informants provided ratings on children's behavior problems for more accurate estimates. Third, we applied models that partition the variance of general versus specific psychopathology. Fourth, we examined the P3 ERP component in early childhood, when neural processes supporting children's executive functions are rapidly developing (Berger, 2011; Diamond, 2002). Fifth, we applied developmental considerations in estimating children's P3 amplitudes. Given rapid brain development from 3–7, and changing P3 amplitudes over this time (van Dinteren et al., 2014), we conducted separate analyses from ages 3–5 and 6–7 to account for these developmental changes. Sixth, we make the data and analysis code publicly available for reproducibility.

Limitations

The study also had weaknesses. First, we examined concurrent associations; thus, we cannot establish causality or the direction(s) of effect. Second, some behavior problem items showed low endorsement rates in the community sample and had to be dropped from analyses, which might weaken our ability to detect associations with the latent factors. Third, fit of the bifactor model was somewhat weaker in terms of one fit index, SRMR. However, SRMR does not account for model parsimony, so better SRMR could be achieved by estimating additional paths that we did not feel would be appropriate according to theory (Greiff & Heene, 2017). By contrast, model fit was strong or acceptable based on indices that account for model parsimony. In addition, relatively few studies have examined bifactor models in early childhood. We are aware of only four samples in which bifactor models of psychopathology have been examined in children as young as three years of age (McElroy et al., 2018; Olino et al., 2010; Sheldrick et al., 2012; Wade et al., 2018). Thus, less is known about general versus specific psychopathology in young children and how best to partition the variance effectively.

Conclusions

The present study is the first to examine the P3 in relation to general and specific psychopathology simultaneously to account for the covariation between externalizing and internalizing problems. Findings suggest that smaller P3 amplitudes may be associated specifically with externalizing problems from a very early age (ages 3–5) and that smaller P3 amplitudes may have implications for a broader range of psychopathology by ages 6–7. The present study highlights the importance of examining neural processes and their associations with general and specific psychopathology across childhood. Findings suggest that there may be a developmental shift in the functional significance of the P3 as it relates to general and specific psychopathology in early childhood.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10802-023-01061-0.

Funding The project was funded by Grants HD098235 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), T32GM108540 from the National Institute of General Medical Sciences (NIGMS), and UL1TR002537 from the National Center for Advancing Translational Sciences (NCATS).

Data Availability Data files, a data dictionary, analysis scripts, and a computational notebook for the present study are published online: https://osf.io/zs2bn. The present study is part of a larger study, the School Readiness Study. Hypotheses and measures for the School Readiness Study were pre-registered: https://osf.io/jzxb8. Hypotheses methods, and a data analysis plan for the present study were also pre-registered: https://osf.io/pny26.

Compliance with Ethical Standards

Ethical Approval The present study was approved by the University of Iowa Institutional Review Board (Study #: 201708761).

Conflict of Interests We have no conflicts of interest to disclose.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont, Research Center for Children, Youth, & Families.

Barry, R. J., De Blasio, F. M., & Fogarty, J. S. (2018). A processing schema for children in the auditory equiprobable Go/NoGo task: ERP components and behaviour. *International Journal of Psychophysiology, 123*, 74–79. https://doi.org/10.1016/j.ijpsycho.2017.10.014

Bechor, M., Ramos, M. L., Crowley, M. J., Silverman, W. K., Pettit, J. W., & Reeb-Sutherland, B. C. (2019). Neural correlates of attentional processing of threat in youth with and without anxiety disorders. *Journal of Abnormal Child Psychology, 47*(1), 119–129. https://doi.org/10.1007/s10802-018-0424-8

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Berger, A. (2011). Self-regulation: Brain, cognition, and development. *American Psychological Association*. https://doi.org/10.1037/12327-000

Bernat, E. M., Ellis, J. S., Bachman, M. D., & Hicks, B. M. (2020). P3 amplitude reductions are associated with shared variance between internalizing and externalizing psychopathology. *Psychophysiology, 57*(7), e13618. https://doi.org/10.1111/psyp.13618

Bitsko, R. H., Claussen, A. H., Lichstein, J., Black, L. I., Jones, S. E., Danielson, M. L., Hoenig, J. M., Jack, S. P. D., Brody, D. J., Gyawali, S., Maenner, M. J., Warner, M., Holland, K. M., Perou, R., Crosby, A. E., Blumberg, S. J., Avenevoli, S., Kaminski, J. W., & Ghandour, R. M. (2022). Mental health surveillance among children—United States, 2013–2019. *Morbidity and Mortality Weekly Report (MMWR) Supplements*, *71*(2), 1–42. https://doi.org/10.15585/mmwr.su7102a1

Boylan, K., Vaillancourt, T., Boyle, M., & Szatmari, P. (2007). Comorbidity of internalizing disorders in children with oppositional defiant disorder. *European Child & Adolescent Psychiatry, 16*(8), 484–494. https://doi.org/10.1007/s00787-007-0624-1

Bruder, G. E., Kayser, J., & Tenke, C. E. (2011). *Event-related brain potentials in depression: Clinical, cognitive, and neurophysiological implications*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0257

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science, 2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Clark, D. A., Hicks, B. M., Angstadt, M., Rutherford, S., Taxali, A., Hyde, L., Weigard, A. S., Heitzeg, M. M., & Sripada, C. (2021). The general factor of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study: A comparison of alternative modeling approaches. *Clinical Psychological Science, 9*(2), 169–182. https://doi.org/10.1177/2167702620959317

Colizzi, M., Lasalvia, A., & Ruggeri, M. (2020). Prevention and early intervention in youth mental health: Is it time for a multidisciplinary and trans-diagnostic model for care? *International Journal of Mental Health Systems, 14*, 23. https://doi.org/10.1186/s13033-020-00356-9

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*(1), 74–101. https://doi.org/10.1037/0033-2909.115.1.74

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195134971.003.0029

Dien, J. (2010). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Dodge, K. A., & Crick, N. R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin, 16*(1), 8–22. https://doi.org/10.1177/0146167290161002

Donchin, E., Miller, G. A., & Farwell, L. A. (1986). The endogenous components of the event-related potential—A diagnostic tool? *Progress in Brain Research, 70*, 87–102. https://doi.org/10.1016/s0079-6123(08)64299-5

Electrical Geodesics, Inc. (2018). *Net Station 5.4 EEG software*.

Feldmann, L., Piechaczek, C. E., Pehl, V., Bartling, J., Bakos, S., Schulte-Körne, G., & Greimel, E. (2018). State or trait? Auditory event-related potentials in adolescents with current and remitted major depression. *Neuropsychologia, 113*, 95–103. https://doi.org/10.1016/j.neuropsychologia.2018.03.035

Forbes, M. K., Tackett, J. L., Markon, K. E., & Krueger, R. F. (2016). Beyond comorbidity Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and Psychopathology, 28*(4pt1), 971–986. https://doi.org/10.1017/S0954579416000651

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25*(4), 355–373. https://doi.org/10.1016/S0149-7634(01)00019-7

Gao, Y., & Raine, A. (2009). P3 event-related potential impairments in antisocial and psychopathic individuals: A meta-analysis. *Biological Psychology, 82*(3), 199–210. https://doi.org/10.1016/j.biopsycho.2009.06.006

Gathercole, S. E. (1998). The development of memory. *Journal of Child Psychology and Psychiatry, 39*(1), 3–27. https://doi.org/10.1111/1469-7610.00301

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment, 33*(5), 313–317. https://doi.org/10.1027/1015-5759/a000450

Greimel, E., Trinkl, M., Bartling, J., Bakos, S., Grossheinrich, N., & Schulte-Körne, G. (2015). Auditory selective attention in adolescents with major depression: An event-related potential study. *Journal of Affective Disorders, 172*, 445–452. https://doi.org/10.1016/j.jad.2014.10.022

Hankin, B. L., Davis, E. P., Snyder, H., Young, J. F., Glynn, L. M., & Sandman, C. A. (2017). Temperament factors and dimensional, latent bifactor models of child psychopathology: Transdiagnostic and specific associations in two youth samples. *Psychiatry Research, 252*, 139–146. https://doi.org/10.1016/j.psychres.2017.02.061

Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child Psychology, 21*(2), 143–150. https://doi.org/10.1207/s15374424jccp2102_6

Hogan, A. M., Butterfield, E. L., Phillips, L., & Hadwin, J. A. (2007). Brain response to unexpected novel noises in children with low and high trait anxiety. *Journal of Cognitive Neuroscience, 19*(1), 25–31. https://doi.org/10.1162/jocn.2007.19.1.25

Hosch, A., Oleson, J. J., Harris, J. L., Goeltz, M. T., Neumann, T., LeBeau, B., Hazeltine, E., & Petersen, I. T. (2022). Studying children's growth in self-regulation using changing measures to account for heterotypic continuity: A Bayesian approach to developmental scaling. *Developmental Science*, *00*, e13280. https://doi.org/10.1111/desc.13280

Houston, R. J., Bauer, L. O., & Hesselbrock, V. M. (2003). Depression and familial risk for substance dependence: A P300 study

of young women. *Psychiatry Research: Neuroimaging, 124*(1), 49–62. https://doi.org/10.1016/S0925-4927(03)00074-X

Hoyniak, C. P., Petersen, I. T., McQuillan, M. E., Staples, A. D., & Bates, J. E. (2015). Less efficient neural processing related to irregular sleep and less sustained attention in toddlers. *Developmental Neuropsychology, 40*(3), 155–166. https://doi.org/10.1080/87565641.2015.1016162

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

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2014). Mental disorders in childhood: Shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA, 311*(17), 1727–1728. https://doi.org/10.1001/jama.2014.1193

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*(2), 199–214. https://doi.org/10.1037/0012-1649.18.2.199

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., & Clark, L. A. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454. https://doi.org/10.1037/abn0000258

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology, 121*(4), 971. https://doi.org/10.1037/a0028355

Lahey, B. B., Moore, T. M., Kaczkurkin, A. N., & Zald, D. H. (2021). Hierarchical models of psychopathology: Empirical support, implications, and remaining issues. *World Psychiatry, 20*(1), 57–63. https://doi.org/10.1002/wps.20824

Lansford, J. E., Malone, P. S., Dodge, K. A., Crozier, J. C., Pettit, G. S., & Bates, J. E. (2006). A 12-year prospective study of patterns of social information processing problems and externalizing behaviors. *Journal of Abnormal Child Psychology, 34*(5), 709–718. https://doi.org/10.1007/s10802-006-9057-4

Lepistö, T., Soininen, M., Čeponiene, R., Almqvist, F., Näätänen, R., & Aronen, E. T. (2004). Auditory event-related potential indices of increased distractibility in children with major depression. *Clinical Neurophysiology, 115*(3), 620–627. https://doi.org/10.1016/j.clinph.2003.10.020

Liu, J., Chen, X., & Lewis, G. (2011). Childhood internalizing behaviour: Analysis and implications. *Journal of Psychiatric and Mental Health Nursing, 18*(10), 884–894. https://doi.org/10.1111/j.1365-2850.2011.01743.x

Lynch, S. J., Sunderland, M., Newton, N. C., & Chapman, C. (2021). A systematic review of transdiagnostic risk and protective factors for general and specific psychopathology in young people. *Clinical Psychology Review*, *87*, 102036. https://doi.org/10.1016/j.cpr.2021.102036

McElroy, E., Belsky, J., Carragher, N., Fearon, P., & Patalay, P. (2018). Developmental stability of general and specific factors of psychopathology from early childhood to adolescence: Dynamic mutualism or p-differentiation? *Journal of Child Psychology and Psychiatry, 59*(6), 667–675. https://doi.org/10.1111/jcpp.12849

Montroy, J. J., Bowles, R. P., Skibbe, L. E., McClelland, M. M., & Morrison, F. J. (2016). The development of self-regulation across early childhood. *Developmental Psychology, 52*(11), 1744–1762. https://doi.org/10.1037/dev0000159

Muthén, L., & Muthén, B. (2021). Mplus user's guide (1998–2021). *Los Angeles, CA: Muthén & Muthén*.

Olino, T. M., Dougherty, L. R., Bufferd, S. J., Carlson, G. A., & Klein, D. N. (2014). Testing models of psychopathology in preschool-aged children using a structured interview-based assessment. *Journal of Abnormal Child Psychology, 42*(7), 1201–1211. https://doi.org/10.1007/s10802-014-9865-x

Olino, T. M., Klein, D. N., Dyson, M. W., Rose, S. A., & Durbin, C. E. (2010). Temperamental emotionality in preschool-aged children and depressive disorders in parents: Associations in a large community sample. *Journal of Abnormal Psychology, 119*(3), 468–478. https://doi.org/10.1037/a0020112

Pasion, R., Fernandes, C., Pereira, M. R., & Barbosa, F. (2018). Antisocial behaviour and psychopathy: Uncovering the externalizing link in the P3 modulation. *Neuroscience & Biobehavioral Reviews, 91*, 170–186. https://doi.org/10.1016/j.neubiorev.2017.03.012

Patrick, C. J., Bernat, E. M., Malone, S. M., Iacono, W. G., Krueger, R. F., & McGue, M. (2006). P300 amplitude as an indicator of externalizing in adolescent males. *Psychophysiology, 43*(1), 84–92. https://doi.org/10.1111/j.1469-8986.2006.00376.x

Pedersen, M. L., Holen, S., Lydersen, S., Martinsen, K., Neumer, S.-P., Adolfsen, F., & Sund, A. M. (2019). School functioning and internalizing problems in young schoolchildren. *BMC Psychology, 7*(1), 88. https://doi.org/10.1186/s40359-019-0365-1

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology, 27*(3), 791–818. https://doi.org/10.1017/S0954579414000789

Petersen, I. T., Hoyniak, C. P., Bates, J. E., Staples, A. D., & Molfese, D. L. (2018). A longitudinal, within-person investigation of the association between the P3 ERP component and externalizing behavior problems in young children. *Journal of Child Psychology and Psychiatry, 59*(10), 1044–1051. https://doi.org/10.1111/jcpp.12975

Pogarell, O., Padberg, F., Karch, S., Segmiller, F., Juckel, G., Mulert, C., Hegerl, U., Tatsch, K., & Koch, W. (2011). Dopaminergic mechanisms of target detection—P300 event related potential and striatal dopamine. *Psychiatry Research: Neuroimaging, 194*(3), 212–218. https://doi.org/10.1016/j.pscychresns.2011.02.002

Polich, J. (2011). Neuropsychology of P300. *Oxford University Press*. https://doi.org/10.1093/oxfordhb/9780195374148.013.0089

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Reeb-Sutherland, B. C., Vanderwert, R. E., Degnan, K. A., Marshall, P. J., Pérez-Edgar, K., Chronis-Tuscano, A., Pine, D. S., & Fox, N. A. (2009). Attention to novelty in behaviorally inhibited adolescents moderates risk for anxiety. *Journal of Child Psychology and Psychiatry, 50*(11), 1365–1372. https://doi.org/10.1111/j.1469-7610.2009.02170.x

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment, 31*, 1395–1411. https://doi.org/10.1037/pas0000754

Santopetro, N. J., Kallen, A. M., Threadgill, A. H., Amir, N., & Hajcak, G. (2022). Blunted flanker P300 demonstrates specificity to depressive symptoms in females during adolescence. *Research on Child and Adolescent Psychopathology, 50*(4), 537–548. https://doi.org/10.1007/s10802-021-00876-z

Sattler, J. M. (2014). *Foundations of behavioral, social, and clinical assessment of children* (6th ed.). Jerome M. Sattler, Publisher, Inc.

Scharf, F., Widmann, A., Bonmassar, C., & Wetzel, N. (2022). A tutorial on the use of temporal principal component analysis in developmental ERP research – Opportunities and challenges. *Developmental Cognitive Neuroscience, 54*, 101072. https://doi.org/10.1016/j.dcn.2022.101072

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance

and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74.
Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Psychology Software Tools Inc.
Schoemaker, K., Mulder, H., Deković, M., & Matthys, W. (2013). Executive functions in preschool children with externalizing behavior problems: A meta-analysis. *Journal of Abnormal Child Psychology, 41*(3), 457–471. https://doi.org/10.1007/s10802-012-9684-x
Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338.
Sheldrick, R. C., Henson, B. S., Merchant, S., Neger, E. N., Murphy, J. M., & Perrin, E. C. (2012). The Preschool Pediatric Symptom Checklist (PPSC): Development and initial validation of a new social/emotional screening instrument. *Academic Pediatrics, 12*(5), 456–467. https://doi.org/10.1016/j.acap.2012.06.008
Trafimow, D. (2015). A defense against the alleged unreliability of difference scores. *Cogent Mathematics, 2*(1), 1064626. https://doi.org/10.1080/23311835.2015.1064626
Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E., & Binau, S. G. (2019). Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of Abnormal Psychology, 128*(3), 185–199. https://doi.org/10.1037/abn0000410
Valadez, E. A., Pine, D. S., Fox, N. A., & Bar-Haim, Y. (2022). Attentional biases in human anxiety. *Neuroscience & Biobehavioral Reviews, 142*, 104917. https://doi.org/10.1016/j.neubiorev.2022.104917
van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: A systematic review and meta-analysis. *PLoS ONE, 9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347
Wade, M., Fox, N. A., Zeanah, C. H., & Nelson, C. A. (2018). Effect of foster care intervention on trajectories of general and specific psychopathology among children with histories of institutional rearing: A randomized clinical trial. *JAMA Psychiatry, 75*(11), 1137–1145. https://doi.org/10.1001/jamapsychiatry.2018.2556
Waldman, I. D., Poore, H. E., van Hulle, C., Rathouz, P. J., & Lahey, B. B. (2016). External validity of a hierarchical dimensional model of child and adolescent psychopathology: Tests using confirmatory factor analyses and multivariate behavior genetic analyses. *Journal of Abnormal Psychology, 125*(8), 1053–1066.
Wang, Y., & Liu, Y. (2021). The development of internalizing and externalizing problems in primary school: Contributions of executive function and social competence. *Child Development, 92*(3), 889–903. https://doi.org/10.1111/cdev.13462
White, J. L., Moffitt, T. E., Earls, F., Robins, L., & Silva, P. A. (1990). How early can we tell?: Predictors of childhood conduct disorder and adolescent delinquency. *Criminology, 28*(4), 507–535. https://doi.org/10.1111/j.1745-9125.1990.tb01337.x
Zhou, Q., Chen, S. H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives, 6*(2), 112–121. https://doi.org/10.1111/j.1750-8606.2011.00176.x

Supplementary Appendix S1. Description of Participants.

Participants were recruited through a biomedical registry of children who had well-child checkups at University of Iowa Hospital and Clinics, university email listservs, and from advertisements and in-person recruitment activities at their school or preschool, Women, Infants, and Children (WIC) programs, pediatricians' offices, and community events. Among participating primary caregivers (n = 125) and parenting partners (n = 118), 97.0% were biological parents, 1.2% were adoptive parents, 0.9% were stepparents, and 1.0% were grandparents. For simplicity, we refer to the primary caregiver and parenting partner as parents. Participants included more primary caregivers than children because the identified primary caregiver changed across time for some children. The composition of marital status among primary caregivers was married (83.7%), remarried (2.3%), separated (0.8%), divorced (4.7%), and single or never married (8.5%). The composition of parents' educational attainment included: completed some high school (1.5%), completed high school (4.9%), completed some college (15.2%), Associate's degree (9.5%), Bachelor's degree (29.7%), Master's degree (22.8%), professional school degree (8.4%), and doctoral degree (8.0%). Compared to the U.S. population, participants in the sample were somewhat more likely to be Non-Hispanic White, married, be middle or upper class, and have a college or graduate degree. Participant demographics were broadly reflective of the surrounding area.

Supplementary Appendix S2. Description of Missing Data.

A total of 94 children had scores for P3 amplitudes; 51 had one time point, 23 had two time points, 19 had three time points, and 1 had four time points. All 124 children had scores for behavior problems. Among possible participant-by-wave instances, 30% had missing scores because the child was not yet eligible for a given wave. Among eligible participant-by-wave instances, approximately 78% had behavior problem ratings. Among missing EEG visits at a given wave for which the child reached eligibility, reasons for missingness included: not interested (14%), too busy (14%), moved/relocated (1%), unable to contact (14%), coronavirus (COVID-19) pandemic (47%), and other (10%). Thus, over half of missing instances were due to the COVID-19 pandemic or to not yet being eligible. We suspended lab visits for 14 months during the COVID-19 pandemic (March 2020 – April 2021). We continued to collect online questionnaires from families during the pandemic but were unable to collect EEG assessments during this period. Among those with EEG visits, reasons for missingness were as follows: child refused to wear the EEG cap (7%), child refused to play the task (2%), not enough good channels (1%), not enough good trials (0%), and another technical problem (2%).

We examined whether missingness was systematic in the predictor (P3) or outcomes (behavior problems). Older children were more likely to be missing P3 (t[214.22] = -3.07, p = .002) and behavior problem (t[7.44] = -3.85, p = .005) scores than younger children, likely due to some COVID-related attrition. In addition, children with more externalizing problems were actually *less* likely to be missing P3 scores than children with fewer externalizing problems (t[397.29] = 2.80, p = .005). Children from lower SES families were more likely to be missing behavior problem scores than children from higher SES families (t[100.40] = 2.55, p = .012). Boys were more likely than girls to be missing behavior problem scores (χ^2[1] = 7.31, p = .007).

Effect sizes of differences were small. Missingness in the P3 was not significantly related to the child's sex, internalizing problems, or the family's socioeconomic status. There was also no difference in missingness as a function of the child's ethnicity, though some frequencies in the cross-tabulation cells were sparse to evaluate this.

Supplementary Appendix S3. Reliability Estimates of ASEBA Measurement of Externalizing and Internalizing Problems.

Mothers' ratings on the Externalizing and Internalizing scales were moderately associated with ratings by fathers (r[139] = .55 and r[139] = .49, respectively, ps < .001). The 9-month stability was strong for mothers' and fathers' ratings of externalizing problems (r[123] = .61 and r[38] = .67, respectively, ps < .001). The 9-month stability of mothers' and fathers' ratings of internalizing problems was moderate-to-strong (r[123] = .67 and r[38] = .46, respectively, ps < .001). Internal consistency was strong for mothers' ratings of externalizing problems for the CBCL 1.5–5 and CBCL 6–18 (α = .91 and .85, respectively), as well as father's ratings (α = .89 and .86, respectively). Internal consistency was strong for mothers' ratings of internalizing problems for the CBCL 1.5–5 and CBCL 6–18 (α = .81 and .78, respectively), and was acceptable for fathers' ratings (α = .75 and .76, respectively).

Supplementary Appendix S4. Collection and Pre-Processing of Electrophysiological Recordings.

Electrophysiological data were collected using an Electrical Geodesic, Inc (EGI) 128-electrode Hydrocel Geodesic Sensor Net with a Net Amps 400 series amplifier. Electrodes were active electrodes composed of silver chloride (Ag-Cl) plated carbon-embedded plastic. The electrode net was soaked in a saline solution before being placed on the child's head. Net Station Acquisition Software 5.4.2 (Electrical Geodesics, Inc., 2018) was used to collect the continuous EEG data. The recording system's precision was .024 μV/bit and had an analog-to-digital conversion rate of 8000 samples per minute. Stimulus presentation was managed using E-Prime 2.0.10.356 (Schneider et al., 2012). Auditory stimuli were presented at a volume of 75 decibels (± 2 decibels). During recording, electrode impedances were adjusted to be at or below 50 kΩ and continuous EEG data were collected at a sampling rate of 1000 Hz.

Data were pre-processed in Net Station Tools 5.4.3 (Electrical Geodesics, Inc., 2018). Continuous data were band-pass filtered from 0.1 to 30.0 Hz, and then segmented into 1200 ms epochs that began 200 ms prior to the presentation of each stimulus. Epochs were then automatically inspected for artifacts, which included identifying and removing "bad" channels. Eye blinks and eye movements were also identified. Channels were marked bad if they contained a voltage shift greater than 200 μV during a given segment length of 80 ms. Eye blinks were classified as a voltage shift greater than 175 μV (max−min) within a 640 ms moving time window for each trial after running a 80 ms moving-average smoothing algorithm across the entire trial period. Eye movements were classified as a voltage shift greater than 200 μV (max−min) over a 640 ms window (with a 80 ms moving-average smoothing algorithm). Epochs were marked bad if they contained more than 20 bad channels, an eye blink, or an eye

movement. Channels were marked bad across all epochs if 20% or greater of the epochs were marked bad. Channels marked bad across all epochs were removed. Removed channels were interpolated based on the waveforms of surrounding electrodes.

If a child did not have at least 10, artifact-free trials in each condition after automatic processing, epochs were manually examined for artifacts. After manually identifying and removing artifacts and bad channels, epochs were subjected to the same automatic inspection procedure described above. Epochs were then averaged within participants, and re-referenced to an average reference (i.e., the average of all scalp electrodes). Finally, epochs were baseline corrected by subtracting the average activity from each epoch's 200 ms baseline. Data were excluded from analyses if the child did not have at least 10 correct, artifact-free trials in each condition after manual processing.

Younger children (ages 3–5) had an average of 20.16 and 45.45 artifact-free infrequent and frequent trials, respectively. Older children (ages 6–7) had an average of 21.69 and 47.46 artifact-free infrequent and frequent trials, respectively. There were no significant differences in the number of infrequent and frequent trials kept per age group (t[149.61] = -1.76, p = .080 and t[140.57] = -1.31, p = .300, respectively). The number of infrequent trials kept was not related to infrequent or frequent P3 amplitudes (r = .01, p = .914 and r = -.03, p = .607, respectively), externalizing problems (r = .04, p = .543), or internalizing problems (r = .01, p = .834). The number of frequent trials kept was not related to infrequent or frequent P3 amplitudes (r = .00, p = .997 and r = .03, p = .617, respectively), externalizing problems (r = .02, p = .723), or internalizing problems (r = .04, p = .542).

Supplementary Appendix S5. Description of Temporospatial PCA.

To perform the sequential temporospatial PCA (tsPCA), we followed recommendations by Dien and colleagues (Dien, 2010, 2012; Dien & Frishkoff, 2005). First, we conducted temporal PCA using a promax rotation to identify distinct temporal components. The number of components to retain in each of the PCA analyses was determined using a parallel test, which compares eigenvalues to eigenvalues from randomly simulated data (Horn, 1965). Then, we conducted spatial PCA using an infomax rotation to identify distinct spatial components. Finally, we conducted spatial PCA using an infomax rotation on the temporal components identified in the previous temporal PCA to identify distinct temporospatial components. The tsPCA was conducted separately for each age group, i.e., 36–54 months and 63–90 months, and separately for each condition, i.e., infrequent and frequent trials.

The temporospatial waveform, thought to correspond with the P3 component, was selected based on a priori hypotheses about the latency, topography, and morphology of the component. Full details on the selected P3 component for each condition are described below.

Ages 36–54 Months

Frequent trials. The temporal PCA retained 23 temporal components, which collectively explained greater than 97 percent of the variance across timepoints in the waveforms. The spatial PCA retained 16 spatial components, which collectively explained greater than 85 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 161 temporospatial components (23 temporal components × 7 spatial components). The selected P3 component explained approximately 0.93 percent of the overall variance in the

waveforms.

Infrequent trials. The temporal PCA retained 24 temporal components, which collectively explained greater than 97 percent of the variance across timepoints in the waveforms. The spatial PCA retained 17 spatial components, which collectively explained greater than 86 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 7 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 168 temporospatial components (24 temporal components × 7 spatial components). The selected P3 component explained approximately 0.66 percent of the overall variance in the waveforms.

Ages 63–90 Months

Frequent trials. The temporal PCA retained 23 temporal components, which collectively explained greater than 97 percent of the variance across timepoints in the waveforms. The spatial PCA retained 17 spatial components, which collectively explained greater than 86 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 6 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 138 temporospatial components (23 temporal components × 6 spatial components). The selected P3 component explained approximately 10.93 percent of the overall variance in the waveforms.

Infrequent trials. The temporal PCA retained 24 temporal components, which collectively explained greater than 97 percent of the variance across timepoints in the waveforms. The spatial PCA retained 17 spatial components, which collectively explained

greater than 86 percent of the variance across electrodes in the waveforms. The spatial PCA on the temporal components retained 6 spatial components, which collectively explained greater than 80 percent of the variance across electrodes in the temporal components. Thus, the two-step temporospatial PCA retained 144 temporospatial components (24 temporal components × 6 spatial components). The selected P3 component explained approximately 8.02 percent of the overall variance in the waveforms.

In sum, the P3 ERP component explained a greater proportion of the variance in the waveforms among the older children compared to the younger children. Age-related differences in the variance accounted for by the P3 could reflect a range of non-mutually exclusive factors, including neurodevelopmental differences, increasing P3 ERP amplitudes across childhood (van Dinteren et al., 2014), and greater signal-to-noise ratio in the older children.

Additional ERP Components

In examining the extracted components, we considered components other than the hypothesized P3 component that also explained variance in the waveform. For example, in both younger (36–54 months) and older (63–90 months) children, a component consistent with the topography and timing of the mismatch negativity was identified (Fitzgerald & Todd, 2020). The mismatch negativity ERP had a frontal distribution peaking from 165–173 ms and explained 2.02 percent of the variance older children, and 1.74 percent of the variance in younger children. However, for theoretical reasons (as described in the text) and to limit multiple testing, we focused on the P3 ERP.

In younger children only, we identified a component characterized by a posterior positivity that explained 7.79 percent of the variance in the infrequent condition, and 11.92 percent in the frequent condition. The component peaked at 373.44 ms for infrequent trials, and

slightly earlier at 321.38 ms for frequent trials. Although this posterior positivity was consistent with the topography and morphology of the P3, the latency was 150 ms earlier than the selected P3 component for the older children in our sample. Research has shown that as children get older, P3 latencies decrease, possibly due to more efficient information processing (van Dinteren et al., 2014). Thus, we did not select the component described above, which peaked earlier (i.e., 321.38 ms – 373.44 ms) than would be expected. The component selected as the P3 for younger children peaked 70 ms later than the component selected for older children, as expected, and was consistent with the expected P3 topography and morphology. Moreover, the selected P3 component appeared to differentiate frequent and infrequent trials, showing (in younger children) a somewhat larger peak for infrequent trials compared to frequent trials, consistent with expectations.

Supplementary Appendix S6. Covariates.

The child's age, child's sex (male = 0, female = 1), family socioeconomic status, informant type (mother = 0, father = 1), the number of infrequent trials kept, and the number of "bad" trials were included as covariates in sensitivity analyses. Socioeconomic status (SES) was calculated as the average of three *z*-scored indices: income-to-needs ratio, parent educational attainment, and parent occupational prestige. Income-to-needs ratio was computed based on the ratio of the household's income relative to poverty thresholds from the U.S. Census Bureau given the number of adults and children in the home. Parent educational attainment was scored as the highest level of education completed: 1 = less than 7^{th} grade; 2 = junior high school; 3 = partial high school; 4 = high school graduate; 5 = partial college (at least one year) or specialized training; 6 = standard college or university graduation; 7 = graduate professional training (graduate degree). Parent occupational prestige was scored using the Nam-Powers-Boyd occupational scale (Boyd & Nam, 2015).

Supplementary Appendix S7. Description of Steps to Fit a Bifactor Model.

First, we dropped items that had no or low variance (i.e., low endorsement rates). Second, if items showed negative loadings on the specific factor (against theoretical expectation), we did not estimate those items' factor loadings on the specific factor. Inclusion of negative factor loadings can result in uninterpretable specific factors (Haywood et al., 2021; Huppert & Fradkin, 2016; Waldman et al., 2015). However, we retained those items for the purposes of estimating the general factor. Then, as necessary, we added correlated residuals to account for residual correlations between items after accounting for the general and specific factors. We made decisions about which correlated residuals to estimate based on the extent of improvement in model fit that would be achieved (i.e., modification indices) in conjunction with theory. For instance, we only estimated correlated residuals within the same construct (i.e., correlations between externalizing items or correlations between internalizing items).

Data were structured in long form such that each child had a row for each combination of timepoint and rater. We used long form to make use of the multiple ratings across raters and timepoints to estimate a bifactor model with additional observations. Dependency in data within the same child (i.e., multiple raters, multiple timepoints) was handled using a cluster variable, which uses a Huber-White sandwich estimator (Huber, 1967; White, 1980) that provides robust standard errors and fit indices to account for non-independence of data. We provide effect sizes of regression parameters using standardized regression coefficients (β), where $\beta < .2$, $.2 < \beta < 0.5$, and $\beta > .5$ are considered small, medium, and large effects, respectively (Acock, 2014).

Supplementary Appendix S8. ECV Calculations.

The explained common variance was calculated by dividing the variance that the general factor explained (i.e., sum of squared general factor loadings) by the total reliable variance (i.e., sum of squared general and specific factor loadings) using Equation 1 (Constantinou, 2019; Rodriguez et al., 2016). Reliable variance is akin to a total variance estimation, but reliable variance does not include an error estimate. Explained common variance for a specific factor (ECVs) was calculated the same way as for the general factor but the numerator was the sum of squared specific factor loadings (e.g., calculating ECVs for externalizing; Equation 2; Constantinou, 2019). ECV and ECVs estimates were calculated separately in the ages 3–5 and ages 6–7 models, because the models had different factor loadings.

$$\frac{\left(\sum \beta^2{}_{\text{General}}\right)}{\left(\sum \beta^2{}_{\text{General}}\right)+\left(\sum \beta^2{}_{\text{Externalizing}}\right)+\left(\sum \beta^2{}_{\text{Internalizing}}\right)} \quad (1)$$

$$\frac{\left(\sum \beta^2{}_{\text{Externalizing}}\right)}{\left(\sum \beta^2{}_{\text{General}}\right)+\left(\sum \beta^2{}_{\text{Externalizing}}\right)+\left(\sum \beta^2{}_{\text{Internalizing}}\right)} \quad (2)$$

Supplementary Appendix S9. Description of Steps to Derive the Final Bifactor Model.

We started by fitting a bifactor measurement model. A bifactor model with all externalizing items (ages 3–5: 24 items; ages 6–7: 35 items) and internalizing items (ages 3–5: 36 items; ages 6–7: 32 items) did not converge. Thus, we followed a stepwise process to derive the final model. First, we dropped items that had little to no variance (i.e., low endorsement rates; ages 3–5: 12 internalizing items; ages 6–7: 19 externalizing items and 19 internalizing items). Many of the items with low endorsement rates at ages 3–5 dealt with somatic complaints (e.g., headaches, stomachaches). Many of the internalizing items with low endorsement rates at ages 6–7 dealt with somatic complaints or suicidal ideation. Many of the externalizing items with low endorsement rates at ages 6–7 dealt with substance use, sexual problems, and low base rate behaviors (e.g., fire setting). Second, we dropped items with negative factor loadings on the specific factor (ages 3–5: 6 externalizing items and 11 internalizing items; ages 6–7: 22 externalizing items and 23 internalizing items). After model adjustments, the general factor was estimated by 48 items for children ages 3–5 and 29 items for children ages 6–7. The externalizing factor was estimated by 18 items for children ages 3–5 and 13 items for children ages 6–7. The internalizing factor was estimated by 13 items for children ages 3–5 and 9 items for children ages 6–7.

The fit of the models was strong according to RMSEA (ages 3–5: .031; ages 6–7: .044), acceptable according to CFI (ages 3–5: .913; ages 6–7: .925) and TLI (ages 3–5: .906; ages 6–7: .915), and somewhat weaker according to SRMR (ages 3–5: .115; ages 6–7: .141). Thus, we added theoretically supported correlated residuals (ages 3–5: 9 correlated residuals; ages 6–7: 2 correlated residuals). After adding correlated residuals, model fit improved slightly according to RMSEA (ages 3–5: .028; ages 6–7: .041), CFI (ages 3–5: .930; ages 6–7: .936), TLI (ages 3–5:

.924; ages 6–7: .927), and SRMR (ages 3–5: .112; ages 6–7: .137).

Supplementary Appendix S10. Sensitivity Analyses.

We conducted several sensitivity analyses, including (a) examining findings with multiply imputed data, (b) examining P3 amplitudes on frequent trials, and (c) examining a P3 amplitude difference score.

Multiple Imputation

We performed multiple imputation as a sensitivity analysis. Given the nonindependence of longitudinal data, we performed multilevel imputation to account for the repeated measurements of data within the same child. Because of the challenges of imputing ordered categorical data, such as the behavior problem ratings (i.e., 0/1/2) in the current study, especially given limited variability in some items, we multiply imputed ERP amplitudes at all time points for which a child had behavior problem ratings. That is, we did not impute scores at all possible time points, because some participants were not yet eligible for some time points due to the ongoing nature of the longitudinal study. Given recommendations (Graham et al., 2007), we performed 100 multiple imputations. We used the informant type, the child's sex, the child's age, the family's socioeconomic status, and the child's behavior problem ratings as predictors in the imputation model. Multilevel imputation due to dependency in data within the same child (i.e., multiple raters, multiple timepoints) was handled using a cluster variable.

There are two broad approaches to multiple imputation: (1) multiple imputation by joint modeling and (2) multiple imputation by chained equations (aka fully conditional specification; (Enders et al., 2016). Multiple imputation by joint modeling assumes that the data follow a joint distribution, most commonly a multivariate normal distribution, which is clearly not the case with our data. Thus, we attempted to perform multiple imputation by chained equations. We performed multiple imputation by chained equations using the mice package (van Buuren &

Groothuis-Oudshoorn, 2011) version 3.15 in R 4.2 (R Core Team, 2022), with a two-level normal model with homogeneous within-group variances using the impute.2l.pan function. However, the multilevel imputation model was unable to be estimated due to singular estimation, which is likely due to the minimal variability in some behavior problem items.

Thus, despite our data not meeting the multivariate normality assumption, we performed multiple imputation via joint modeling. We performed multiple imputation via Mplus version 8.6 (Muthén & Muthén, 2021). We then fit the analysis model for each age group in Mplus.

In the ages 3–5 model, P3 amplitudes on infrequent trials were significantly associated with unique externalizing problems ($\beta = -.22$, $p = .025$), but were not significantly associated with unique internalizing problems or the general factor. Thus, the multiple imputation findings were similar to the model without multiple imputation.

In the ages 6–7 model, P3 amplitudes on infrequent trials were not significantly associated with unique externalizing problems, unique internalizing problems, or the general factor. The effect size of the association between P3 amplitudes and the general factor ($\beta = -.37$) was similar to the effect size in the model without multiple imputation, but the standard errors were somewhat wider. The wider standard errors yielded an estimate that was not statistically significant. This may be due, in part, to the somewhat smaller sample in the ages 6–7 model.

Nevertheless, we present the findings from multiple imputation with caution because our data do not meet the multivariate assumption of normality of the multiple imputation model we applied. Moreover, researchers have argued against using multiple imputation with longitudinal data due to the unreliability of results (Twisk et al., 2013).

P3 Amplitudes on Frequent Trials

As an additional sensitivity analysis, we examined P3 amplitudes on frequent trials in

relation to general versus specific psychopathology in the ages 3–5 bifactor model. P3 amplitudes on frequent trials were not significantly associated with unique externalizing problems, unique internalizing problems, or the general factor.

We also examined P3 amplitudes on frequent trials in relation to general versus specific psychopathology in the ages 6–7 bifactor model. P3 amplitudes on frequent trials were not significantly associated with unique externalizing problems or the general factor. However, P3 amplitudes on frequent trials were positively associated with unique internalizing problems (β = 0.37, p = .049).

P3 Amplitude Difference Score

We also examined a P3 amplitude difference score as a sensitivity analysis. A child's P3 amplitude difference score was calculated as their P3 amplitude on infrequent trials minus their P3 amplitude on frequent trials.

In the ages 3–5 model, P3 amplitude difference scores were not significantly associated with unique externalizing problems or unique internalizing problems. However, P3 amplitude difference scores were negatively associated with the general factor (β = -.17, p = .034).

In the ages 6–7 model, P3 amplitude difference scores were not significantly associated with unique externalizing problems, unique internalizing problems, or the general factor. Nevertheless, we present the difference score findings with caution. Difference scores are notorious for having low reliability. The reliability of difference scores tends to be much lower than the individual indices when (a) the individual indices are correlated (as was the case in the present study; r[254] = .39, p < .001; Revelle & Condon, 2019) and when (b) the individual indices have similar variances (as was the case in the present study; P3 amplitudes on infrequent trials: σ^2 = 3.53; P3 amplitudes on frequent trials: σ^2 = 3.18; Trafimow, 2015).

Supplementary Table S1.

Standardized Factor Loadings for Ages 3–5 Bifactor Model.

	General Factor			Externalizing			Internalizing		
Indicator	**β**	**SE**	***p***	**β**	**SE**	***p***	**β**	**SE**	***p***
Item 2	0.52	0.07	< .001						
Item 4	0.51	0.06	< .001						
Item 5	0.67	0.06	< .001						
Item 6	0.67	0.05	< .001						
Item 7	0.48	0.07	< .001				0.39	0.08	< .001
Item 8	0.62	0.05	< .001						
Item 10	0.55	0.07	< .001						
Item 15	0.58	0.05	< .001	0.64	0.05	< .001			
Item 16	0.62	0.05	< .001	0.31	0.08	< .001			
Item 18	0.69	0.06	< .001	0.31	0.09	.001			
Item 20	0.46	0.07	< .001	0.78	0.06	< .001			
Item 21	0.49	0.07	< .001				0.43	0.08	< .001
Item 23	0.51	0.07	< .001						
Item 24	0.43	0.07	< .001						
Item 27	0.55	0.06	< .001	0.32	0.07	< .001			
Item 29	0.72	0.04	< .001	0.17	0.07	.012			
Item 33	0.36	0.08	< .001				0.29	0.10	.004
Item 35	0.54	0.09	< .001	0.30	0.10	.002			
Item 37	0.61	0.07	< .001						
Item 40	0.54	0.07	< .001	0.41	0.08	< .001			
Item 42	0.44	0.11	< .001	0.19	0.13	.144			
Item 43	0.41	0.13	.001				0.63	0.09	< .001
Item 44	0.78	0.04	< .001	0.14	0.09	.117			
Item 46	0.52	0.19	.005						
Item 47	0.56	0.09	< .001				0.54	0.09	< .001
Item 51	0.97	0.06	< .001				0.21	0.10	.040
Item 53	0.63	0.09	< .001	0.26	0.14	.059			
Item 56	0.32	0.09	< .001						
Item 58	0.67	0.04	< .001	0.29	0.07	< .001			
Item 59	0.58	0.05	< .001						
Item 62	0.65	0.09	< .001						
Item 66	0.77	0.05	< .001	0.16	0.09	.075			
Item 68	0.32	0.08	< .001				0.36	0.11	.001
Item 69	0.55	0.05	< .001	0.31	0.08	< .001			
Item 70	0.65	0.11	< .001						

(Continued)

Table S1 Continued

Item 79	0.64	0.07	< .001				0.39	0.09	< .001
Item 81	0.64	0.05	< .001	0.36	0.07	< .001			
Item 82	0.75	0.05	< .001						
Item 83	0.41	0.11	< .001				0.14	0.12	.259
Item 85	0.76	0.04	< .001	0.32	0.06	< .001			
Item 86	0.29	0.10	.003				0.74	0.11	< .001
Item 87	0.49	0.09	< .001				0.59	0.08	< .001
Item 88	0.66	0.06	< .001	0.51	0.06	< .001			
Item 95	0.69	0.08	< .001	0.22	0.09	.017			
Item 96	0.67	0.04	< .001						
Item 97	0.56	0.06	< .001						
Item 98	0.53	0.13	< .001				0.48	0.11	< .001
Item 99	0.35	0.11	.001				0.67	0.07	< .001

Model Fit	**CFI**	**TLI**	**RMESA**	**SRMR**	***df***
	.930	.924	.028	.112	1047

Note. Items derived from the Child Behavior Checklist (CBCL) 1.5–5; β = standardized factor loading; Standard Error (SE) of standardized factor loadings. “CFI” = comparative fit index; “TLI” = Tucker Lewis index; “RMSEA” = root mean square error of approximation; “SRMR” = standardized root mean square residual; “*df*” = degrees of freedom.

Supplementary Table S2.

Standardized Factor Loadings for the Ages 6–7 Bifactor Model.

	General Factor			Externalizing			Internalizing		
Indicator	**β**	**SE**	***p***	**β**	**SE**	***p***	**β**	**SE**	***p***
Item 3	0.32	0.11	.005	0.79	0.07	< .001			
Item 14	0.67	0.08	< .001						
Item 19	0.66	0.10	< .001	0.17	0.13	.174			
Item 20	0.60	0.14	< .001	0.31	0.14	.025			
Item 21	0.56	0.16	.001	0.31	0.15	.037			
Item 22	0.46	0.12	< .001	0.88	0.10	< .001			
Item 26	0.56	0.12	< .001	0.41	0.15	.006			
Item 28	0.23	0.12	.070	0.80	0.07	< .001			
Item 29	0.26	0.12	.023				0.34	0.14	.018
Item 32	0.33	0.12	.008				0.60	0.11	< .001
Item 35	0.54	0.15	< .001				0.64	0.10	< .001
Item 42	0.70	0.15	< .001						
Item 43	0.39	0.11	< .001	0.79	0.08	< .001			
Item 45	0.38	0.15	.013				0.85	0.13	< .001
Item 47	0.39	0.12	.001						
Item 49	0.56	0.09	< .001				0.19	0.16	.253
Item 50	0.65	0.07	< .001				0.51	0.15	< .001
Item 63	0.52	0.10	< .001						
Item 68	0.95	0.05	< .001	0.21	0.12	.074			
Item 69	0.83	0.14	< .001						
Item 71	0.37	0.12	.001				0.51	0.12	< .001
Item 75	0.44	0.12	< .001				0.29	0.15	.054
Item 81	0.70	0.13	< .001	0.61	0.13	< .001			
Item 86	0.56	0.10	< .001	0.47	0.11	< .001			
Item 87	0.65	0.07	< .001	0.42	0.10	< .001			
Item 88	0.46	0.10	< .001						
Item 95	0.77	0.06	< .001	0.35	0.09	< .001			
Item 104	0.80	0.08	< .001						
Item 112	0.32	0.11	.004				0.68	0.12	< .001
Model Fit	**CFI**	**TLI**		**RMESA**	**SRMR**		***df***		
	.936	.927		.041	.137		353		

Note. Items derived from the Child Behavior Checklist (CBCL) 6–18. β = standardized factor loadings; Standard Error (SE) of standardized factor loadings. “CFI” = comparative fit index;

“TLI” = Tucker Lewis index; “RMSEA” = root mean square error of approximation; “SRMR” = standardized root mean square residual; “*df*” = degrees of freedom.

Supplementary Table S3.

Regression Coefficients of Predictors and Covariates for Ages 3–5 Model

		General Factor				Externalizing				Internalizing			
Model	**Parameter**	**B**	**β**	**SE**	***p***	**B**	**β**	**SE**	***p***	**B**	**β**	**SE**	***p***
No Covariates	P3	-0.02	-0.06	0.03	.454	-0.07	-0.24	0.03	.015	0.03	0.11	0.04	.348
With Covariates	P3	-0.02	-0.06	0.03	.489	-0.12	-0.34	0.04	.004	0.08	0.21	0.06	.183
With Covariates	Age	-0.02	-0.02	0.08	.759	-0.19	-0.14	0.11	.078	0.41	0.26	0.14	.003
With Covariates	Sex	-0.19	-0.09	0.16	.239	-0.90	-0.38	0.27	.001	0.99	0.38	0.34	.003
With Covariates	SES	-0.12	-0.09	0.11	.190	0.13	0.08	0.21	.557	-0.23	-0.13	0.26	.384
With Covariates	Role	0.29	0.14	0.11	.008	-0.13	-0.06	0.16	.415	-0.70	-0.27	0.23	.002
With Covariates	Trials Kept	0.01	0.04	0.01	.622	0.04	0.21	0.03	.074	-0.06	-0.27	0.03	.070
With Covariates	Bad Channels	0.00	0.05	0.00	.479	-0.01	-0.13	0.01	.152	0.02	0.27	0.01	.023

Note. Age in years. Sex is coded such that 1 = female and 0 = male. “SES” = socioeconomic status. “Trials Kept” is the final number of trials not excluded during the oddball P3 task. “Bad Channels” is the count of channels excluded from analysis due to poor quality. “P3” is the child’s P3 amplitude on the infrequent trials of the oddball task. “Externalizing” and “Internalizing” represent the latent factors for specific externalizing and internalizing problems, respectively. Two models were fit separately: the P3 without covariates (“No Covariates”) and the P3 with covariates (“With Covariates”).

Supplementary Table S4.

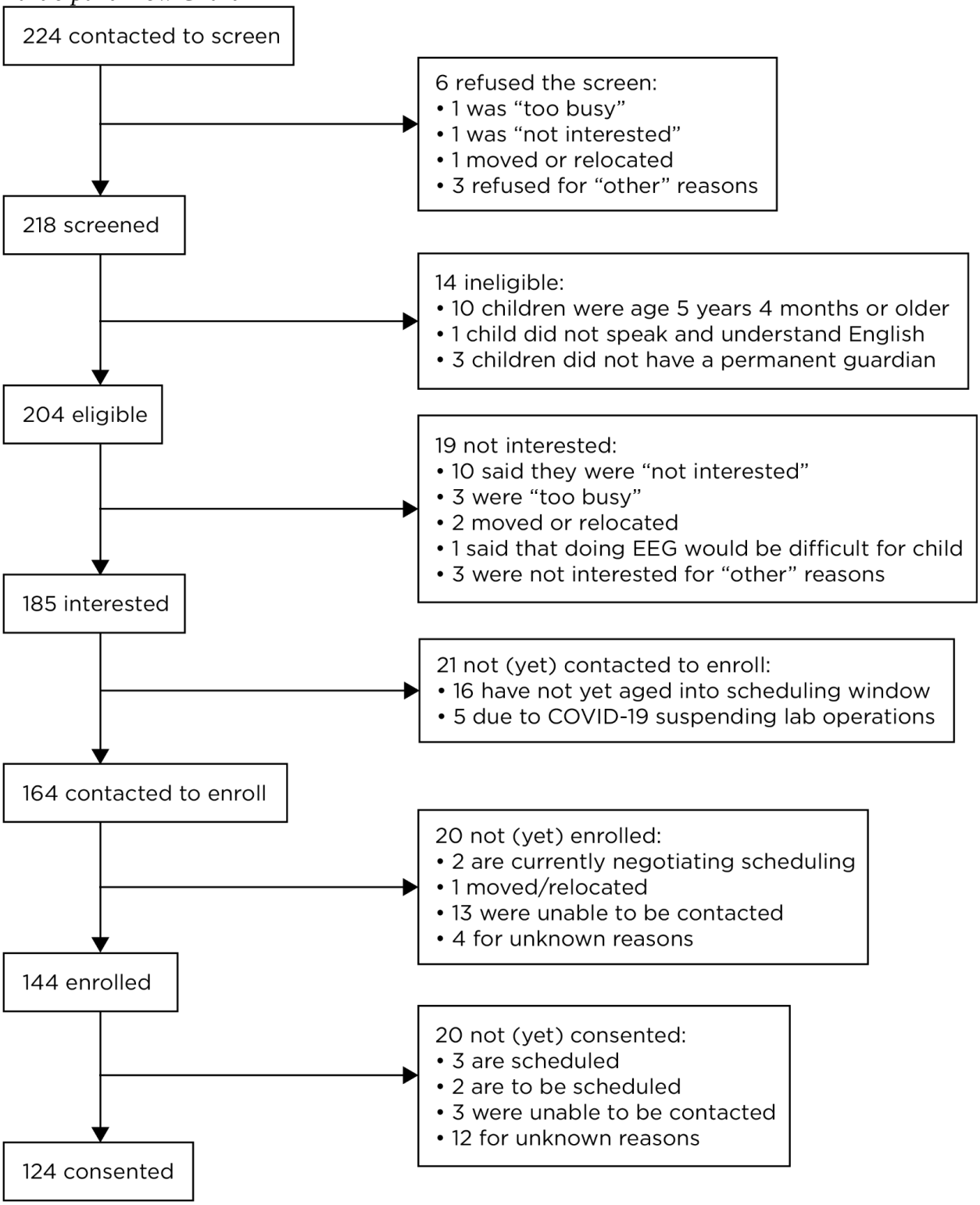
Regression Coefficients of Predictors for Ages 6–7 Model

	General Factor				Externalizing				Internalizing			
Parameter	**B**	**β**	**SE**	***p***	**B**	**β**	**SE**	***p***	**B**	**β**	**SE**	***p***
P3	-0.11	-0.39	0.05	.018	-0.04	-0.16	0.05	.384	0.14	0.47	0.10	.189

Note. "P3" is the child's P3 amplitude on the infrequent trials of the oddball task. "Externalizing" and "Internalizing" represent the latent factors for specific externalizing and internalizing problems, respectively. A model with all covariates did not converge.

Supplementary Figure S1

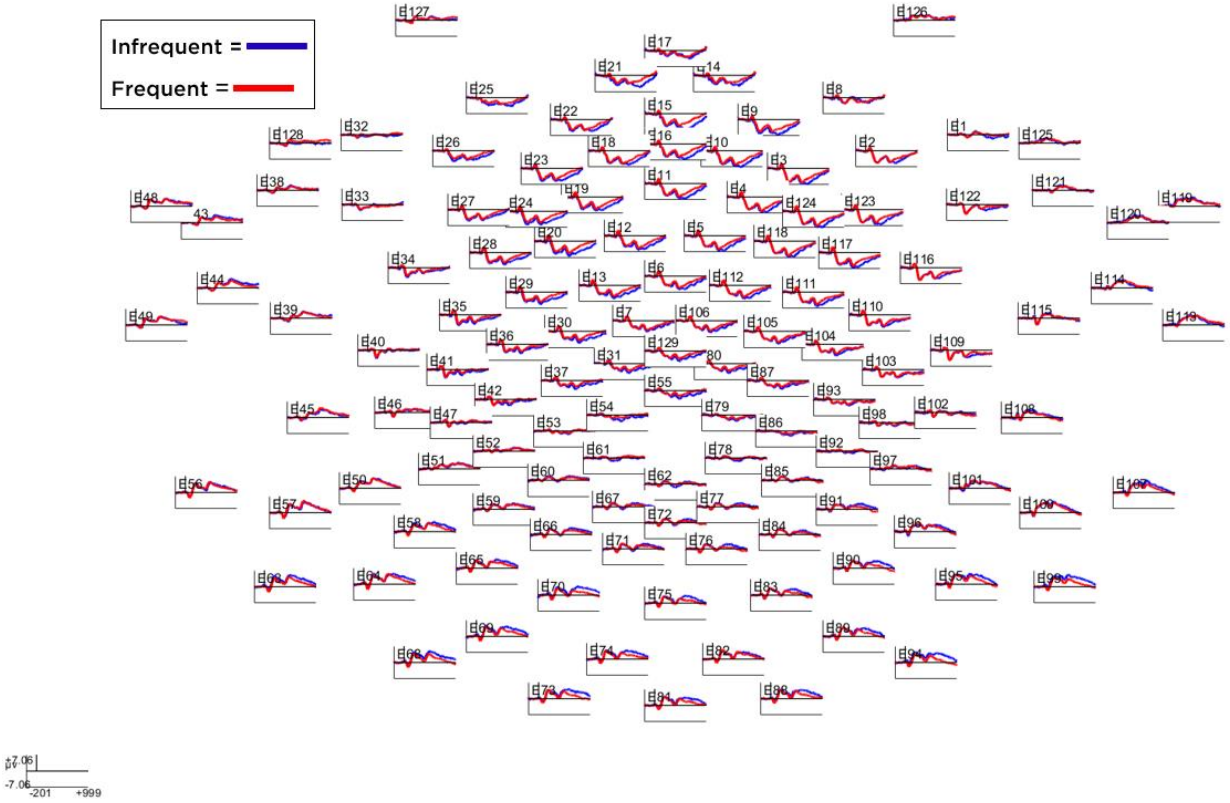
Participant Flow Chart



Note. EEG = electroencephalography

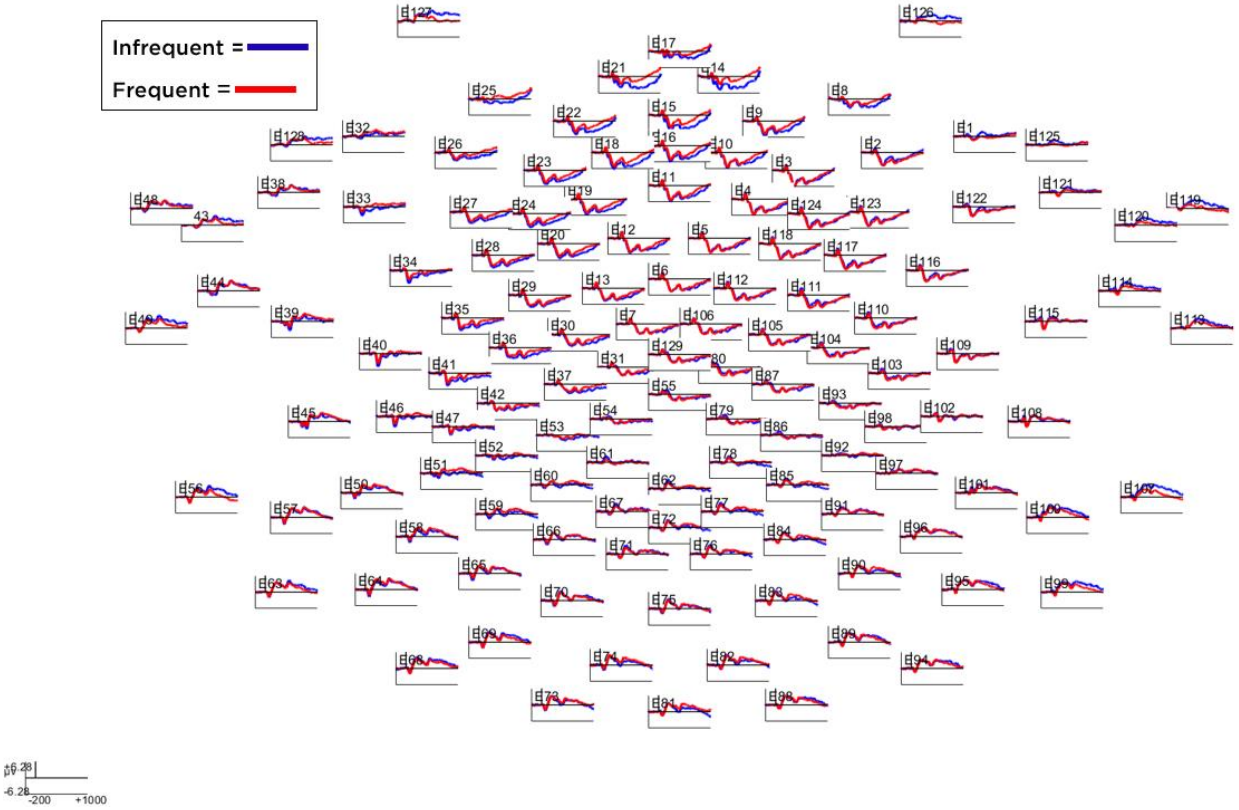
Supplementary Figure S2.

Topo Plot: Grand-Averaged Waveforms Ages 3–4



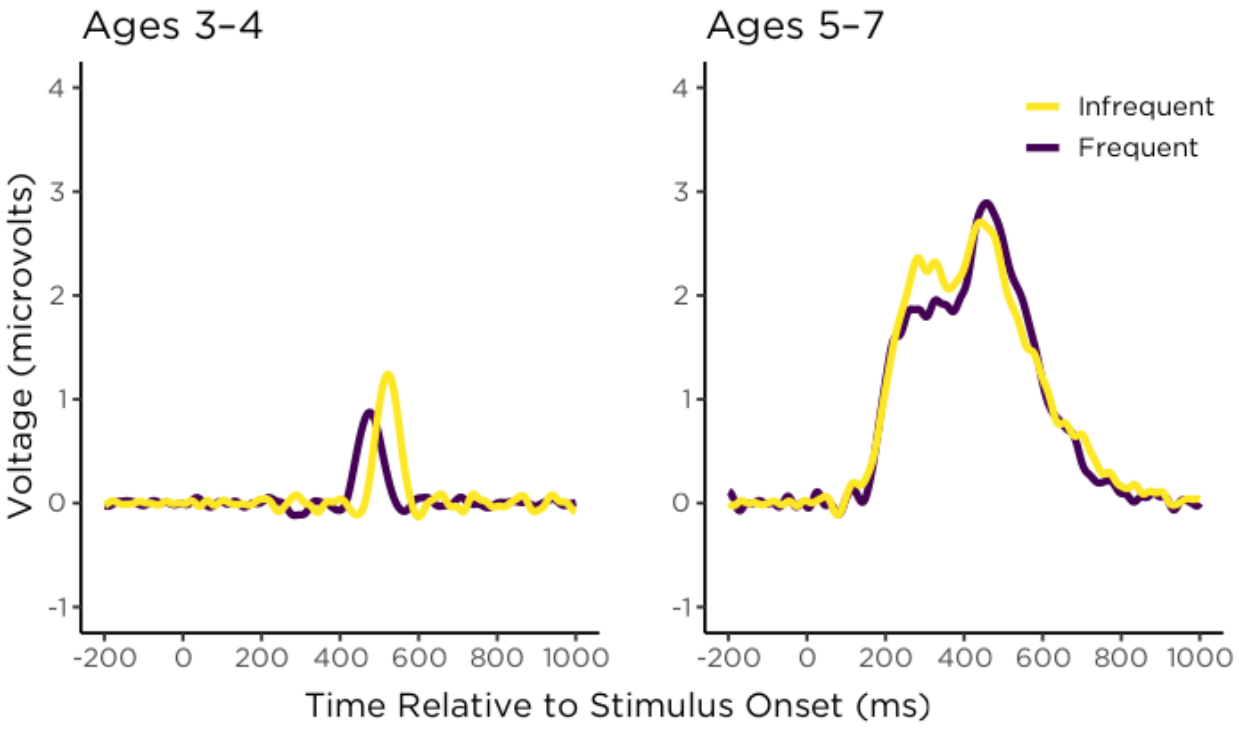
Supplementary Figure S3.

Topo Plot: Grand-Averaged Waveforms Ages 5–7



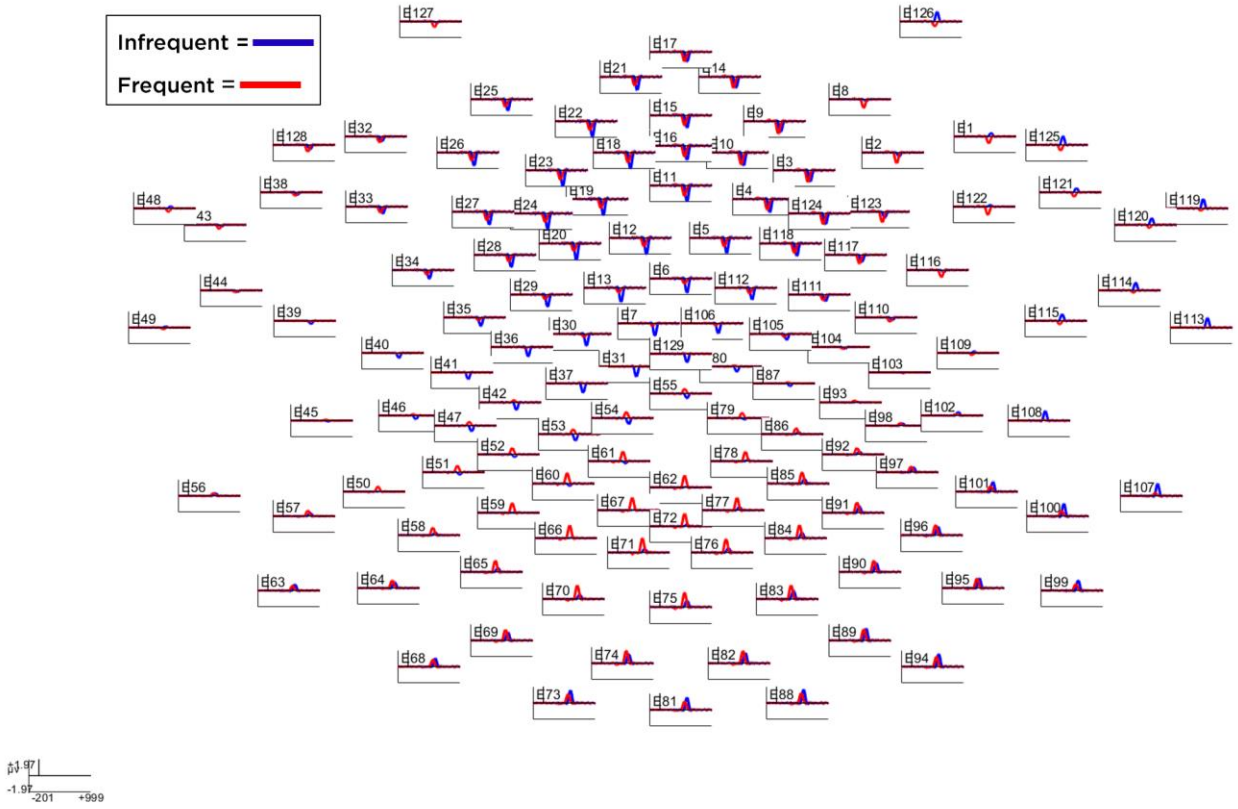
Supplementary Figure S4.

P3 ERP PCA Component Waveform



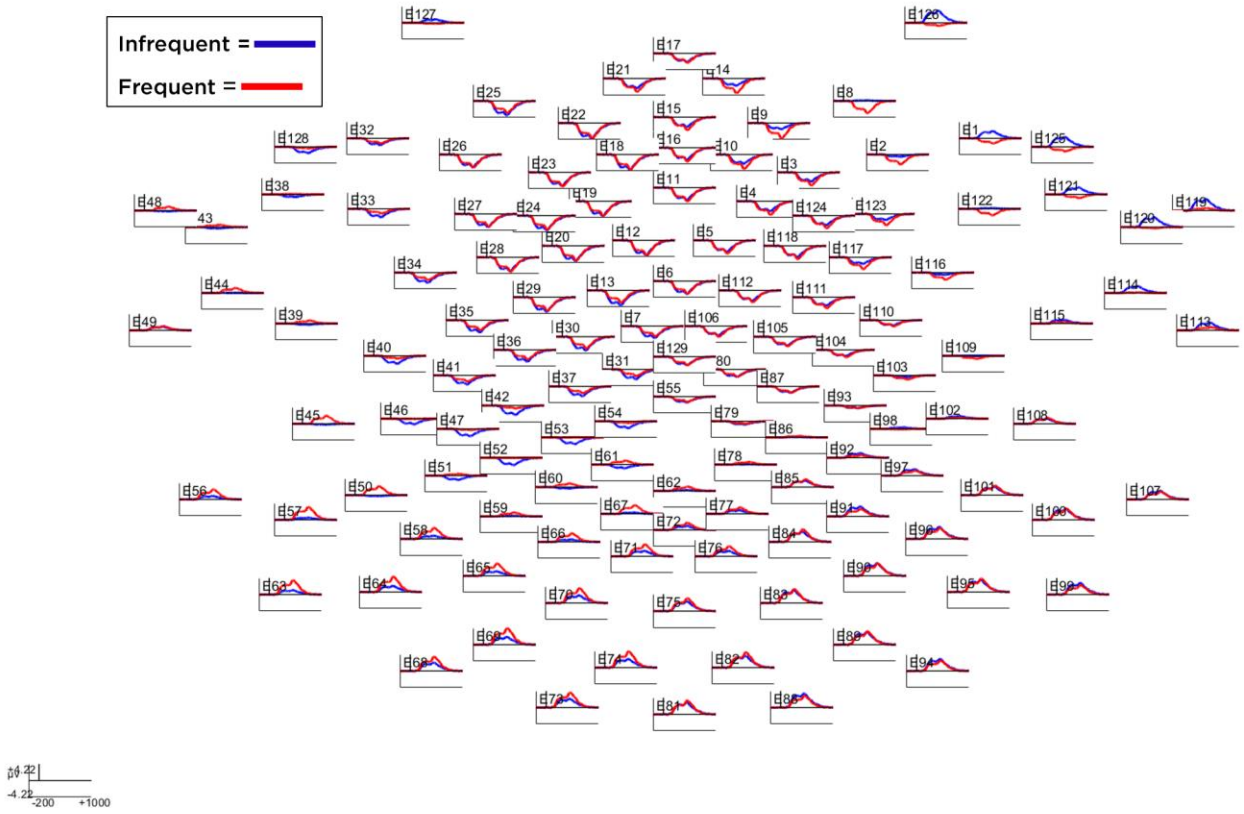
Supplementary Figure S5.

Topo Plot: P3 ERP PCA Component Waveforms Ages 3–4



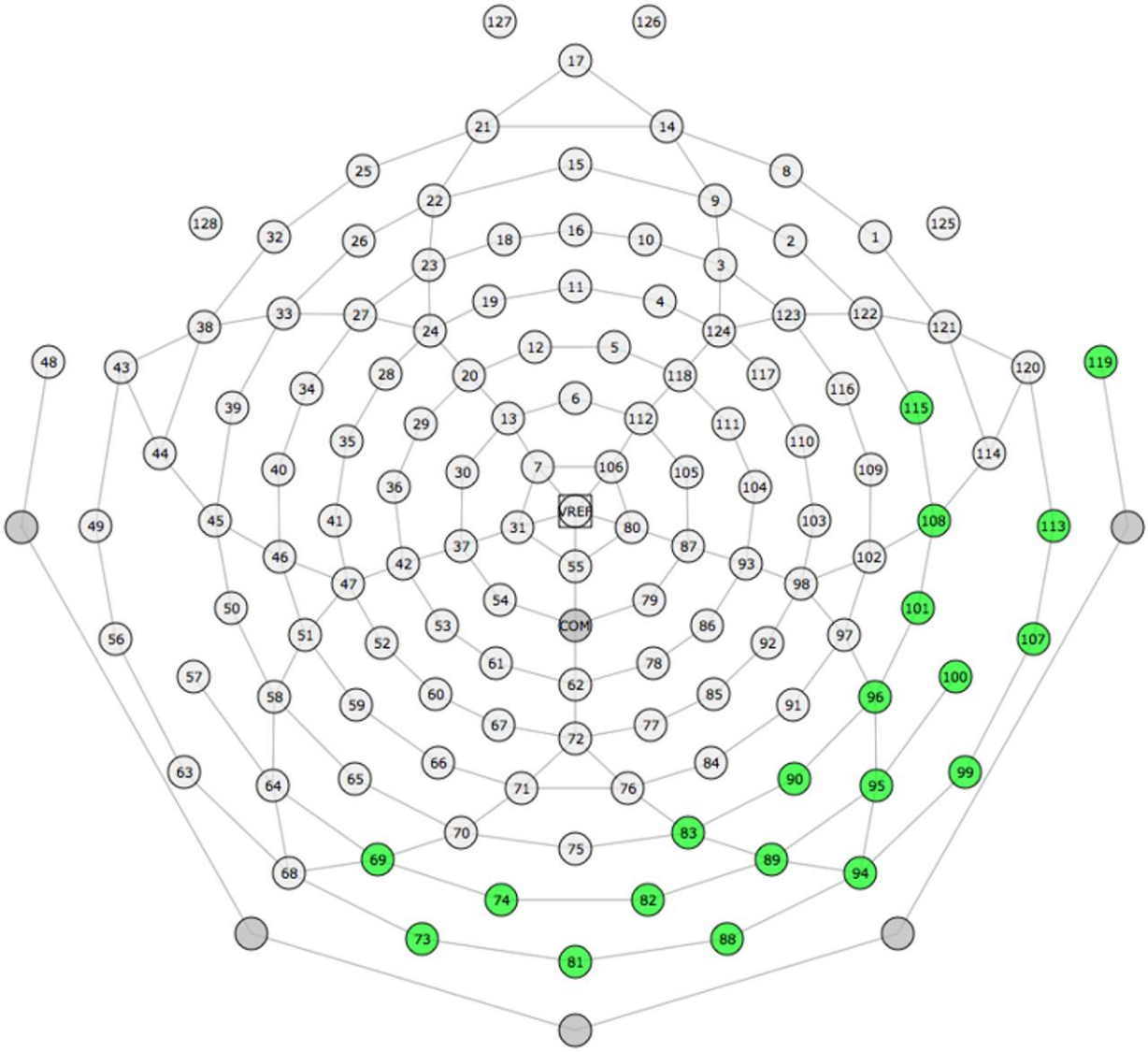
Supplementary Figure S6.

Topo Plot: P3 ERP PCA Component Waveforms Ages 5–7



Supplementary Figure S7.

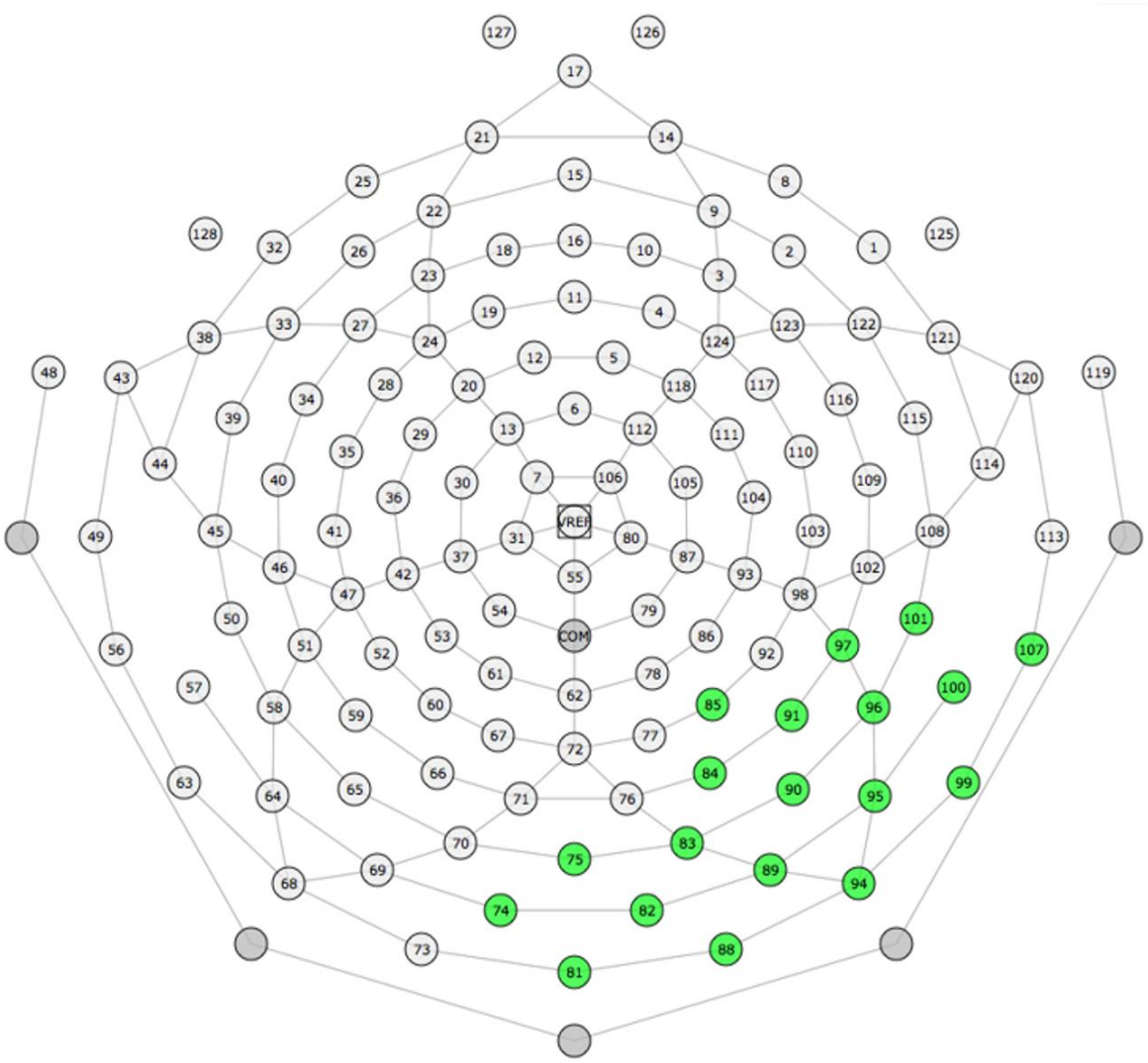
P3 ERP Electrode Cluster Ages 3–4



Note. Green electrodes correspond to electrodes whose loading on the P3 temporospatial component was .45 or greater.

Supplementary Figure S8.

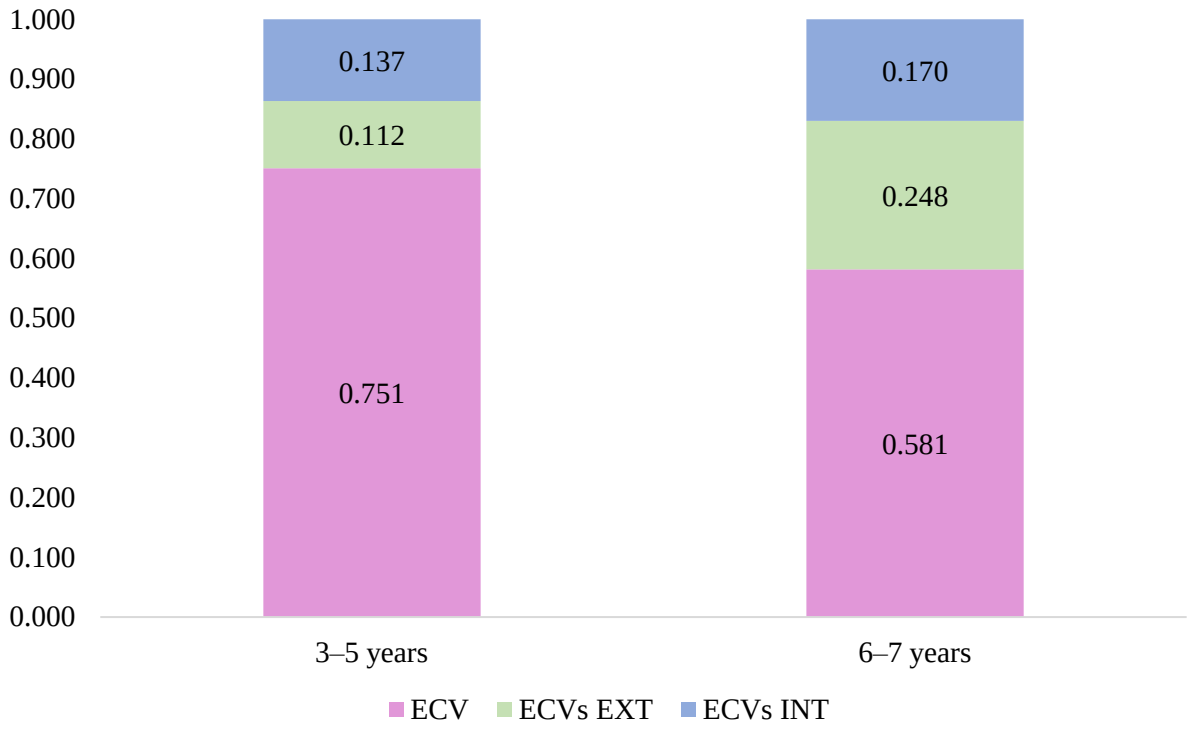
P3 ERP Electrode Cluster Ages 5–7



Note. Green electrodes correspond to electrodes whose loading on the P3 temporospatial component was .45 or greater.

Supplementary Figure S9.

Explained Common Variance (ECV) and Explained Common Variance of Specific Factor (ECVs) estimates



Note. "ECV" = explained common variance. "ECVs" = explained common variance of specific factor. "EXT" = externalizing. "INT" = internalizing. ECV and ECVs are reported as proportion of sum of squared factor loadings of factor divided by total reliable variance, i.e., sum of squared factor loadings of general and specific factor loadings. Reliable variance for the ages 3–5 model was 21.917, and for the ages 6–7 model was 16.001. ECV, ECVs EXT, and ECVs INT are ordered from bottom to top in the figure.

References

Acock, A. C. (2014). *A gentle introduction to Stata* (4th ed.). Stata Press.

Boyd, M., & Nam, C. B. (2015). *The newest nam-powers-boyd occupational scale: Development and insights*. Southern Demographic Association annual meeting.

Constantinou, M. (2019). *The Bifactor Model of Psychopathology: Methodological Issues and Clinical Applications*.

Dien, J. (2010). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*(1), 170–183. https://doi.org/10.1111/j.1469-8986.2009.00885.x

Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Developmental Neuropsychology*, *37*(6), 497–517. https://doi.org/10.1080/87565641.2012.697503

Dien, J., & Frishkoff, G. A. (2005). Introduction to principal components analysis of event-related potentials. In T. C. Handy (Ed.), *Event related potentials: A methods handbook* (pp. 189–207). MIT Press.

Electrical Geodesics, Inc. (2018). *Net Station 5.4 EEG software*.

Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation. *Psychological Methods*, *21*, 222–240. https://doi.org/10.1037/met0000063

Fitzgerald, K., & Todd, J. (2020). Making sense of mismatch negativity. *Frontiers in Psychiatry*, *11*, 468. https://doi.org/10.3389/fpsyt.2020.00468

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213. https://doi.org/10.1007/s11121-007-0070-9

Haywood, D., Baughman, F. D., Mullan, B. A., & Heslop, K. R. (2021). One p-factor for all? Exploring the applicability of structural models of psychopathology within subgroups of a population. *International Journal of Environmental Research and Public Health*, *18*(13). https://doi.org/10.3390/ijerph18137108

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. University of California Press.

Huppert, J. D., & Fradkin, I. (2016). Validation of the Penn Inventory of Scrupulosity (PIOS) in scrupulous and nonscrupulous patients: Revision of factor structure and psychometrics. *Psychological Assessment*, *28*(6), 639–651. https://doi.org/10.1037/pas0000203

Muthén, L., & Muthén, B. (2021). Mplus user's guide (1998–2021). *Los Angeles, CA: Muthén & Muthén*.

R Core Team. (2022). *R: A language and environment for statistical computing*.

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*, 1395–1411. https://doi.org/10.1037/pas0000754

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.

Trafimow, D. (2015). A defense against the alleged unreliability of difference scores. *Cogent Mathematics*, *2*(1), 1064626. https://doi.org/10.1080/23311835.2015.1064626

Twisk, J., de Boer, M., de Vente, W., & Heymans, M. (2013). Multiple imputation of missing values was not necessary before performing a longitudinal mixed-model analysis. *Journal of Clinical Epidemiology*, *66*(9), 1022–1028. https://doi.org/10.1016/j.jclinepi.2013.03.017

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

van Dinteren, R., Arns, M., Jongsma, M. L. A., & Kessels, R. P. C. (2014). P300 development across the lifespan: A systematic review and meta-analysis. *PLOS ONE*, *9*(2), e87347. https://doi.org/10.1371/journal.pone.0087347

Waldman, I., Rathouz, P., Van Hulle, C., & Lahey, B. (2015). Testing alternative models for the underlying structure of childhood psychopathology using confirmatory factor analyses and multivariate behavior genetic analyses. *Behavior Genetics*, *45*(6), 692.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934