The Journal of Child Psychology and Psychiatry

Journal of Child Psychology and Psychiatry 59:10 (2018), pp 1044–1051 doi:10.1111/jcpp.12975

A longitudinal, within-person investigation of the association between the P3 ERP component and externalizing behavior problems in young children

Isaac T. Petersen,[1] Caroline P. Hoyniak,[2] John E. Bates,[2] Angela D. Staples,[3] and Dennis L. Molfese[4]

[1]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA; [2]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA; [3]Department of Psychology, Eastern Michigan University, Ypsilanti, MI, USA; [4]Department of Psychology, University of Nebraska-Lincoln, Lincoln, NE, USA

Background: Externalizing problems, including aggression and conduct problems, are thought to involve impaired attentional capacities. Previous research suggests that the P3 event-related potential (ERP) component is an index of attentional processing, and diminished P3 amplitudes to infrequent stimuli have been shown to be associated with externalizing problems and attention-deficit/hyperactivity disorder (ADHD). However, the vast majority of this prior work has been cross-sectional and has not examined young children. The present study is the first investigation of whether *within*-individual changes in P3 amplitude predict changes in externalizing problems, providing a stronger test of developmental process. **Method:** Participants included a community sample of children ($N = 153$) followed longitudinally at 30, 36, and 42 months of age. Children completed an oddball task while ERP data were recorded. Parents rated their children's aggression and ADHD symptoms. **Results:** Children's within-individual changes in the P3 amplitude predicted concomitant within-child changes in their aggression such that smaller P3 amplitudes (relative to a child's own mean) were associated with more aggression symptoms. However, changes in P3 amplitudes were not significantly associated with ADHD symptoms. **Conclusions:** Findings suggest that the P3 may play a role in development of aggression, but do not support the notion that the P3 plays a role in development of early ADHD symptoms. **Keywords:** P3 ERP; externalizing behavior problems; aggression; attention-deficit/hyperactivity disorder; early childhood.

Introduction

Externalizing behaviors in early childhood predict maladaptive outcomes in adulthood, including substance use and criminality (Petersen, Bates, Dodge, Lansford, & Pettit, 2015). One key feature of externalizing disorders is impaired attentional capacities. Event-related potentials (ERPs), stimulus-locked neuro-electrical activity measured using electroencephalography (EEG), can be used to study neural correlates of attentional processing that mark impaired attention in externalizing problems. The present study advances understanding of developmental process in externalizing problems by examining the longitudinal association between neural functioning and externalizing problems.

P3 ERP and attention processing

The oddball task, in which two stimuli are presented, one frequent and the other infrequent, is commonly used to assess neural correlates of attention. A widely studied index of response to infrequent stimuli in the oddball task, the P3 ERP component, the third positive deflection in the waveform poststimulus, is considered an index of attention. The predominant theory of the P3 is that it indexes attention and memory processes, reflecting neural mechanisms associated with updating mental representations stored in working memory based on novel incoming stimuli (Polich, 2012). The cognitive interpretation of the P3 depends on the task in which it is elicited. In the current study's passive oddball task, in which no behavioral response was required, the P3 likely represents attentional orienting (Friedman, Cycowicz, & Gaeta, 2001). Attentional orienting involves rapid, passive attentional shifts to new/unexpected stimuli, and may reflect early evaluation of stimulus importance to determine whether further cognitive processing is necessary (Hermens et al., 2010).

A robust literature has examined the P3 cross-sectionally at different points in childhood. Meta-analytic evidence suggests that the auditory P3 amplitude increases across childhood, plateauing at age 20, whereas its latency decreases across the life span (van Dinteren, Arns, Jongsma, & Kessels, 2014). Additionally, there are slight variations in the electrode regions where the P3 is maximal across development (Hoyniak, Petersen, McQuillan, Staples, & Bates, 2015; Johnstone, Barry, Anderson, & Coyle, 1996). Despite differences in morphology and topography of the P3 elicited in children, an established literature suggests that the P3 is an index of attentional processing in childhood (Hoyniak et al., 2015; Johnstone, Barry, & Clarke, 2013; van Dinteren et al., 2014).

Conflicts of interest statement: No conflicts declared.

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

P3 and externalizing problems

Attentional orienting deficits play a key role in the social information processing style of individuals with high levels of aggression (e.g., failing to notice cues of nonaggressive intent; Dodge & Crick, 1990), explaining why diminished attention orienting capacities might be associated with externalizing problems. Therefore, the P3 may tap increased risk for externalizing problems. Given robust findings demonstrating an association between the P3 and externalizing problems, the P3 may be an endophenotype (i.e., an intermediate phenotype) of externalizing psychopathology, reflecting the biological processes underlying externalizing problems (Iacono & Malone, 2011). Meta-analyses have shown that smaller P3 amplitudes are associated with externalizing problems (Gao & Raine, 2009; Pasion, Fernandes, Pereira, & Barbosa, 2018). The P3 shows rank-order stability across adulthood (Yoon, Malone, & Iacono, 2015), and has been shown to predict later criminality (Gao, Raine, Venables, & Mednick, 2013). Individual differences in the P3, which are highly heritable, may be a marker of genetic risk for externalizing psychopathology (van Beijsterveldt & van Baal, 2002), and the association between the P3 and externalizing problems is considered genetically mediated (Hicks et al., 2007). Additionally, the P3 is generated by dopaminergic neurotransmission (Pogarell et al., 2011) and by a distributed neural circuit including the lateral prefrontal cortex (Polich, 2007; Soltani & Knight, 2000), both of which have shown impairments in externalizing disorders (Gatzke-Kopp et al., 2009).

Despite considerable research linking the P3 to externalizing psychopathology, the vast majority of this prior work focused on adults and adolescents. We are aware of no studies that have examined the P3 in relation to externalizing problems in toddlers and preschoolers, so it is unclear whether diminished P3 amplitudes might play a role in development of externalizing problems in early childhood. Early childhood is an ideal window to study development of the P3 in relation to externalizing problems because (a) early childhood is characterized by rapid neurodevelopment supporting attention and self-regulatory processes (Diamond, 2002), and (b) externalizing behaviors are common in early childhood, especially physical aggression, which reaches its highest level during this era (Tremblay, 2002). Individual differences in aggression are highly stable and appear as early as toddlerhood (Olweus, 1979). Better understanding of early neural processes associated with development of externalizing problems may lead to earlier, more precisely targeted prevention efforts.

Prior research on the P3 across childhood has mostly been cross-sectional and has not examined whether within-individual changes in the P3 predict within-individual changes in externalizing problems. Investigating the association within the individual removes between-subject confounds by using the individual as their own control, which is a stronger test of causality than between-subjects approaches (Duckworth, Tsukayama, & May, 2010). Based on prior work, we cannot ascertain whether the P3 plays a role in development of externalizing problems or whether it is simply a marker of risk for externalizing psychopathology in general. The P3 has associations with a number of disordered phenotypes, including aggression (Patrick, 2008), attention-deficit/hyperactivity disorder (ADHD; Tsai, Hung, & Lu, 2012), depression, and schizophrenia (Turetsky et al., 2015).

The present study

Longitudinal studies of the association between the P3 and externalizing problems in early childhood can elucidate neural processes associated with development of externalizing behaviors. The present study is the first, to our knowledge, to examine the *within*-individual association between the P3 and externalizing problems. Different associations can be observed at the group level (between-individual) and at the individual level (within-individual), and mistakenly attributing a between-individual association to a within-individual association is known as the ecological fallacy (Curran & Bauer, 2011). The present study examined the longitudinal association between a neural index of attentional processing, the P3, and externalizing behavior problems in early childhood, examining both within- and between-individual associations. Children were followed longitudinally at 30, 36, and 42 months of age. Our main question was whether *within*-child changes in P3 amplitude predicted *within*-child changes in parent-reported externalizing behavior, examining aggression and ADHD symptoms separately.

Method

Participants

A community sample of children and their families ($N = 182$) were recruited from the Bloomington, Indiana area to participate when the children were 30, 36, and 42 months of age. Sample characteristics are reported in Table 1. Participants were assessed within 1 month of their target age. Children completed an oddball ERP task, and primary caregivers (97% mothers) reported on the child's behavior problems. To be included for analysis in the present study, children had to provide usable EEG data in the oddball task ($N = 165$, 91%). Exclusion criteria included nonfebrile seizures ($n = 2$), head injury ($n = 9$), and psychotropic medication ($n = 2$—none of which were psychostimulants), resulting in a final sample of 153.

Due to planned missingness and censoring (i.e., children not yet age-eligible), in the current sample, a total of 375 EEG assessments were possible. Of these 375 possible assessments, 73 assessments were not scheduled because the parent elected not to schedule a visit for their child (65 assessments) or equipment malfunctioning prevented us from collecting EEG data at that time (eight assessments). Hence, 302 EEG assessments were scheduled. Of 302 scheduled assessments,

Table 1 Sample characteristics

Variable	*n*	%
Sex		
Males	83	54
Females	70	46
All	153	
Parent ethnicity		
Non-Hispanic caucasian	138	90
Hispanic	4	3
African-American	5	3
Asian-American	5	3
Mixed race	1	<1
Parent education		
Some high school	1	<1
GED	2	1
High school diploma	2	1
Some college	16	11
College degree	131	86
Marital status		
Single	10	7
Married	135	89
Divorced	6	4

Variable	*M*	*SD*
Child age (months)	36.00	4.90
Parent age (years)	33.26	4.85
Family SES	48.99	13.27

One child was missing information on her parent's education, and two were missing parental marital status. Family socioeconomic status (SES) was calculated using the Hollingshead (1975) index.

257 provided usable EEG data. These 257 assessments were provided by 153 different children ($n_{1\ time}$ = 74, $n_{2\ times}$ = 54, $n_{3\ \ times}$ = 25). Further details on missingness are in Appendix S1 and Table S1.

Measures

Externalizing behavior problems. Externalizing behavior problems were measured using mother report on the Child Behavior Checklist (CBCL 1½–5; Achenbach & Rescorla, 2000). We examined the two subscales comprising the Externalizing scale: Aggression (19 items) and Attention Problems (five items). The Aggression subscale includes items about physical aggression, destruction, anger, noncompliance, and attention demands. The Attention Problems subscale has been interpreted as a measure of ADHD symptoms because it assesses the three dimensions of ADHD symptoms: inattention, hyperactivity, and impulsivity (Lifford, Harold, & Thapar, 2008). It is associated with other measures of ADHD, including the Conners Rating Scale and DSM-IV symptoms of ADHD (Derks et al., 2008). In addition, it has been shown to measure ADHD as accurately as the Conners Rating Scale (Derks et al., 2008), with strong sensitivity and specificity (Chen, Faraone, Biederman, & Tsuang, 1994).

The CBCL is among the best normed and most widely used measures for behavior problems in this age range, has good test–retest reliability and good validity (content, criterion, construct; Sattler & Hoge, 2006). Primary caregivers rated whether a behavior was *not true* (0), *somewhat or sometimes true* (1), or *very or often true* (2), and scores were summed across items, with higher levels reflecting more behavior problems. This continuous approach to scoring behavior problems is consistent with evidence that externalizing problems and ADHD are dimensional not categorical (Coghill & Sonuga-Barke, 2012). For this sample, eight children's scores were above the borderline clinical threshold of a *T*-score ≥65 (i.e. above ≥1.5 *SD*) for externalizing problems. Cronbach's alpha in this sample was .90 for aggression and .69 for ADHD symptoms. Cross-time continuity was r = .63 for aggression and r = .60 for ADHD symptoms (df = 189, ps < .001). Children who had scores for externalizing problems were $n_{30\ months}$ = 146, $n_{36\ months}$ = 120, $n_{42\ months}$ = 101.

P3 ERP. Children participated in an oddball task and a go/no-go task while EEG data were collected during a lab visit. The present study focuses on the P3 ERP from the oddball task. A 6-minute auditory oddball (two-tone discrimination) task was used to elicit a P3 ERP component to infrequent sounds. The task was passive; children were not instructed to respond to any stimuli. Although the P3 from passive and active tasks have different latencies, spatial distributions, and cognitive interpretations (Polich, 2007), they both index attentional processing, and smaller amplitudes of both kinds of P3 have been associated with externalizing disorders (Rydkjær et al., 2017; Tsai et al., 2012). We chose to use a passive oddball task because more trials of usable data would be available than a task that required a behavioral response. ERP measures in a passive task may be especially useful in early childhood when behavioral response capacities are still developing and thus less stable, which would complicate interpretation of behavioral task performance.

Pure, low-frequency (1,000 Hz) and high-frequency (1,500 Hz) tones were randomly presented so that one tone occurred on 70% of trials (84 trials; frequent stimulus) and the other tone occurred on 30% of trials (36 trials; infrequent stimulus), with 120 trials in total. The two tones were counterbalanced as frequent versus infrequent across children. Each tone lasted for approximately 300 ms, and the task included an interstimulus interval that varied randomly from 2,300–2,500 ms to prevent habituation. Children were not asked to make a behavioral response. During presentation of the auditory tones, children watched a child-friendly cartoon video on a monitor with the video's audio turned off. On average, participants contributed 25.30 (SD = 7.67) usable infrequent and 58.22 (SD = 18.25) frequent trials.

Netstation Acquisition software version 4.4.2 (Electrical Geodesics, Inc., Eugene, OR) was used to collect and process EEG data from a 128-electrode Hydrocel Geodesic Sensor Net with a sampling rate of 250 Hz. Before recording began, electrode impedances were adjusted lower than 50 kΩ. Children's continuous EEG data were band-pass filtered from 0.3 to 30 Hz, and epochs 1,200 ms in duration were extracted, beginning 200 ms prior to the presentation of the target stimulus. Data were then visually inspected for artifacts. Following visual inspection, a channel was marked as bad if a voltage change of greater than 150 μV occurred during a given segment of length 80 ms, and a segment was marked as bad if it contained 20 or more bad channels. On average, participants had 10.07 (SD = 3.83) bad channels of 128 channels.

Epoched data were then re-referenced to the average reference (i.e., subtracting the average potential of all channels from the potential at each channel), and baseline corrected by subtracting the average activity over the 200 ms baseline period representing grand-averaged waveforms (Figure 1). After processing, primary components of the ERP waveform were statistically decomposed using a sequential temporospatial principal components analysis (PCA), which objectively and empirically determines regions of electrodes and time frames that parsimoniously account for the variance in the waveforms, and whose components correspond to ERP components (Dien & Frishkoff, 2005). Children's P3 amplitudes were calculated as their mean amplitudes for the temporospatial component reflecting the P3 based on timing,

morphology, and spatial topography. We provide more information about the temporo-spatial PCA in Appendix S2. PCA-derived P3 waveforms are depicted in Figure S1. Cross-time continuity of the P3 amplitude was $r[87] = .21$ ($p = .043$, two-tailed), suggesting some rank-order stability but also considerable neurodevelopmental change from 30 to 42 months of age.

Statistical analysis

Using hierarchical linear modeling (HLM), which handles missingness and unbalanced data (Singer & Willett, 2003), we fit growth curve models with random intercepts and slopes to each child's trajectory of externalizing problems. Growth curve models examined whether within-child changes in P3 amplitudes predicted concomitant within-child changes in externalizing problems, controlling for between-child associations of P3 amplitudes with externalizing problems. Models included the child's (a) mean P3 amplitude across time in association with the child's intercept of externalizing problems (i.e., their level of externalizing problems at 30 months of age): γ_{02}, (b) mean P3 amplitude across time in association with the child's linear slope of externalizing problems (i.e., their change in externalizing problems from 30 to 42 months of age): γ_{12}, (c) time-varying P3 amplitude (centered around the child's mean P3 amplitude across occasions) predicting concomitant within-child changes in externalizing problems: γ_{20}, and (d) covariates that could plausibly account for the association between P3 amplitudes and externalizing problems, including the child's sex as a time-invariant covariate to account for the well-established sex differences in externalizing problems, and time-varying covariates to account for potentially systematic ERP missingness (number of bad channels and number of infrequent trials kept). Centering the P3 amplitude around the child's own mean (so-called person-mean centering) follows best practice for disaggregating within- and between-individual effects (Curran & Bauer, 2011). An assumption of the disaggregation is that the person-level mean P3 amplitude is estimated without error. We were interested in both within-individual (γ_{20}) and between-individual (γ_{02}, γ_{12}) effects. The between-child effects examined whether children's mean P3 amplitudes across time were associated with their intercepts or slopes of externalizing problems. The within-child effect examined whether individuals' time-specific deviations in the P3 amplitude away from their own mean predicted their time-specific deviations in externalizing problems *over and above* their linear slopes (i.e., β_{1i}) of externalizing problems. Thus, the within-child effect examined whether within-child changes in the P3 amplitude predicted concomitant within-child changes in externalizing problems. We fit separate growth curve models for aggression and ADHD symptoms. Model equations and information about the models are in Appendix S3. As a sensitivity analysis, we examined models with multiple imputation (Appendix S3). Because the substantive findings were unchanged (Table S4), results from the raw data are presented.

Results

Descriptive statistics and correlations between study variables are in Tables S2 and S3. Bivariate correlations showed that P3 amplitudes were negatively associated with aggression but not significantly correlated with ADHD symptoms (Table S3).

Next, we examined the within- and between-child associations between P3 amplitudes and externalizing problems using HLM growth curve models. HLM growth curve model results are in Table 2. The child's mean P3 amplitude across time was negatively associated with their intercepts (γ_{02}: $B = -0.30$, $p = .039$), but not their slopes, of aggression. Within-individual changes in the P3 amplitude were negatively associated with within-child changes in aggression (γ_{20}: $B = -0.31$, $p = .004$). Findings suggest that smaller P3 amplitudes (relative to one's mean) were concurrently associated with more aggression (relative to one's level of aggression at other time points) above and beyond one's linear slope of aggression. Findings held even accounting for ADHD symptoms (Table S5). The child's mean P3 amplitude across time was not significantly associated with their intercepts (γ_{02}) or slopes (γ_{12}) of ADHD symptoms. Within-child changes in P3 amplitudes were not significantly associated with within-child changes in ADHD symptoms (γ_{20}: $B = -0.06$, $p = .12$).

Discussion

The present study examined the longitudinal, within-person association between P3 ERP amplitudes and parent-reported externalizing problems in very young children. Our findings suggest that smaller mean P3 amplitudes across ages 30–42 months were associated with higher levels of aggression at 30 months of age. Additionally, findings suggest that within-child changes in the P3 amplitude were negatively associated with concomitant within-child changes in aggression. When children showed smaller P3 amplitudes (relative to their own mean level), they showed more concurrent aggression. However, within-child changes in P3 amplitudes were not significantly associated with ADHD symptoms.

Our findings of an association between smaller P3 amplitudes and aggression are consistent with prior meta-analyses examining externalizing problems (Gao & Raine, 2009; Pasion et al., 2018) and with conceptualizations of the P3 as an endophenotype of externalizing problems (Iacono & Malone, 2011). The auditory P3 amplitude increases across childhood (van Dinteren et al., 2014), and is considered an index of attentional orienting (Friedman et al., 2001).

Although the P3 amplitude is often found to be smaller in children with ADHD, as compared to controls (e.g., Tsai et al., 2012), contradictory findings have also emerged (e.g., Rydkjær et al., 2017). Yoon, Iacono, Malone, Bernat, and McGue (2008) found that children with ADHD and a comorbid externalizing disorder (oppositional defiant disorder or conduct disorder) had smaller P3 amplitudes, whereas children with ADHD alone did not show such an effect. These findings suggest that the smaller P3 amplitude typically noted in children with ADHD might actually reflect comorbid externalizing problems. Moreover, to our knowledge, no studies that identified an association between P3 amplitudes and ADHD examined whether within-individual changes in P3 were associated with ADHD. Our findings align

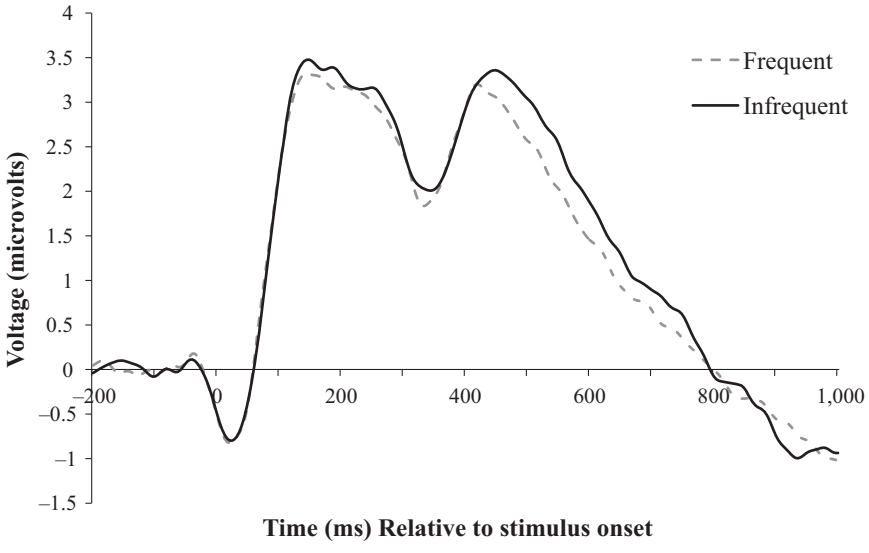


Figure 1 Children's grand-averaged waveforms for frequent and infrequent trials during the oddball task. The waveform depicted represents the mean waveform from those electrodes with a 0.6 or greater factor loading onto the PCA component reflecting the P3. For purposes of depicting waveforms, electrodes from the PCA-derived posterior/parietal electrode cluster (see Figure S1) were averaged with equal, unit weighting. However, actual P3 amplitudes were calculated using PCA

with findings of prior studies that the smaller P3 amplitude in children with externalizing problems is not due to ADHD symptoms (Baving, Rellum, Laucht, & Schmidt, 2006). Thus, evidence does not support a causal interpretation of an association between the P3 and ADHD; prior findings of an association between the P3 and ADHD could reflect their common association with a third variable (e.g., broad factor of externalizing problems). This finding could reflect an important developmental fact about the meaning of the P3 in the oddball task, or less within-individual variation in ADHD symptoms than in aggression across 30–42 months (Appendix S3). Or perhaps attention problems are a less coherent or stable construct in very early development. Or the finding could reflect measurement issues such as slightly weaker reliability of the shorter ADHD scale compared to the longer aggression scale.

By contrast, the within-individual association between changes in P3 amplitude and aggression provides stronger evidence consistent with a causal association, even though we cannot eliminate the possibility of time-varying confounds or the reverse direction of effect. How might a smaller P3 amplitude be involved in development of aggression? First, it is important to note that a smaller P3 amplitude may reflect not just underprocessing of relevant information, but also overprocessing of irrelevant information (Hermens et al., 2010), which could impair higher order processes related to detecting and responding to subtle environmental and social cues. Within a social information processing framework, underprocessing of relevant information (e.g., cues of safety) and overprocessing of irrelevant information (e.g., ambiguous cues perceived as indicating threat) hypothetically could affect the first stage of social information processing, encoding of social cues. Altered encoding could, in turn, influence downstream attributions, making it more likely that individuals interpret ambiguous cues as hostile, and respond with aggression. Within a social information processing framework, impaired attention and encoding processes could explain, at a basic stimulus processing level, why individuals with smaller P3 amplitudes show more aggression, particularly reactive (as opposed to proactive) aggression. This is consistent with findings showing smaller P3 amplitudes in impulsive aggression, but not premeditated aggression (for a review, see Patrick, 2008). For instance, children with poorer novelty detection (e.g., smaller P3 amplitudes), may miss key changes in others' voice tone in daily interactions (Hoyniak et al., 2018), leading to agonistic conflicts with others, which in turn lead to future assumptions about hostile intent and negatively biased social information processing.

The present study had several key strengths. First, we examined the P3 ERP in very young children, an important group with high theoretical relevance for understanding how externalizing problems develop. Based on the morphology and topography of the P3, evidence suggests the P3 component elicited in the present study may correspond to the P3 elicited from older subjects. Our findings contribute to a relatively sparse literature focusing on ERPs elicited during toddlerhood. Second, the study was longitudinal with repeated measures of both the P3 and behavior problems. The repeated measures design allowed us to examine whether within-child changes in P3 amplitude predicted concomitant within-child changes in behavior problems. We believe this is the first study to examine the within-individual association between the P3 and externalizing problems, providing a stronger test of causality.

Table 2 Results of HLM growth curve models

Outcome	*B*	*β*	*SE*	*df*	*p*
Aggression					
Intercept	**15.52**	**0.01**	**2.27**	**142**	**<.001**
Time	−0.11	0.00	0.13	90	.371
Sex	−1.07	−0.07	1.00	142	.288
Sex × Time	0.06	0.03	0.12	90	.607
Mean P3 amplitude	**−0.30**	**−0.13**	**0.15**	**142**	**.039**
Mean P3 amplitude × Time	0.02	0.05	0.02	90	.370
***Time-varying P3 amplitude**	**−0.31**	**−0.12**	**0.10**	**90**	**.004**
*Number of bad channels	−0.09	−0.06	0.09	90	.306
***Number of infrequent trials kept**	**−0.13**	**−0.14**	**0.06**	**90**	**.022**
ADHD symptoms					
Intercept	**3.30**	**0.03**	**0.75**	**142**	**<.001**
Time	−0.06	−0.07	0.04	90	.126
Sex	−0.66	−0.08	0.35	142	.064
Sex × Time	0.08	0.11	0.04	90	.053
Mean P3 amplitude	−0.03	−0.05	0.05	142	.538
Mean P3 amplitude × Time	0.00	0.01	0.01	90	.878
*Time-varying P3 amplitude	−0.06	−0.07	0.04	90	.122
*Number of bad channels	−0.01	−0.01	0.03	90	.849
*Number of infrequent trials kept	−0.01	−0.04	0.02	90	.546

'Time' reflects the slope term, and is centered at the first time point so that the intercept reflects the child's level at 30 months (i.e. 0, 6, 12 months from 30 months). 'Sex' is coded with female = 1, male = 0. 'Mean P3 amplitude' refers to a given child's mean P3 amplitude across time (time invariant). 'Time-varying P3 amplitude' refers to a given child's P3 amplitude at a given time point that is centered around their mean P3 amplitude across time (time varying). Interaction terms with time essentially reflect the prediction of slopes of the outcome (e.g. 'Sex × Time' reflects sex predicting slopes of the outcome). Asterisks reflect time-varying terms. Terms in bold reflect significant associations at $p < .05$ level.

The present study also had limitations. First, because of the correlational nature of the design, and the many likely determinants of psychological development, we cannot make definitive causal inferences. Next, our sample was predominantly middle class, which may limit generalizability of our findings. We hope to see future studies with more broadly representative as well as higher risk samples and the use of informants beyond parents. The extent of longitudinal ERP missingness warrants caution in interpreting our findings. Despite its clinical non-specificity, the P3 may have transdiagnostic relevance. An interesting further question would be how the P3 relates in early childhood to additional dimensions of behavior problems, including internalizing and thought-disordered problems, given findings that the P3 is also associated with depression and schizophrenia (Turetsky et al., 2015).

Conclusion

The present longitudinal study is the first investigation of the *within*-individual association between the P3 and externalizing problems, which provides a stronger test of causality than previous studies of between-subjects effects. Findings indicate that children's within-individual changes in the P3 amplitude predicted concomitant within-child changes in their aggression but not ADHD symptoms. Importantly, this association was present in toddlerhood, an era when early targeted intervention efforts may efficiently prevent later, severe externalizing problems. Findings are consistent with the notion that the P3 may play a role in development of aggression. They are not consistent with the P3 playing a causal role in development of ADHD.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Missingness.
Appendix S2. Temporo-spatial principal components analysis.
Appendix S3. Information about the HLM growth curve models.
Table S1. EEG observations per child by number of assessments and age at time of visit.
Table S2. Descriptive statistics and correlation matrix of P3 amplitude, aggression, and ADHD symptoms at 30, 36, and 42 months of age.
Table S3. Descriptive statistics and correlation matrix of model variables.
Table S4. Results of HLM growth curve models with multiple imputation.
Table S5. Results of HLM growth curve model of aggression while accounting for ADHD symptoms (and additional covariates).
Figure S1. Left panel: Children's PCA-derived P3 waveforms/Right panel: Topo plot of the PCA component reflecting the P3.

Acknowledgements

The Toddler Development Study was funded by Grant HD073202 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. IP was supported by a National Research Service Award from the National Institute of Mental Health (1F31MH100814-01A1). The authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Isaac T. Petersen, Department of Psychological and Brain Sciences, University of Iowa, 309 Stuit Hall, Iowa City, IA 52242, USA; E-mail: isaac-t-petersen@uiowa.edu

Key points

- Previous research suggests that externalizing problems are characterized by attentional impairments and smaller P3 amplitudes.
- Prior work has been mainly cross-sectional, has not examined young children, and has not examined whether *within*-individual changes in P3 amplitude predict changes in externalizing behavior.
- Our findings indicated that children's within-individual changes in the P3 amplitude predicted concomitant within-child changes in their aggression but not ADHD symptoms.
- Findings support the interpretation that the P3 may play a role in development of aggression. They do not support such a role in development of ADHD.

References

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms and Profiles: An integrated system of multi-informant assessment*. Burlington, VT: Department of Psychiatry, University of Vermont.

Baving, L., Rellum, T., Laucht, M., & Schmidt, M.H. (2006). Children with oppositional-defiant disorder display deviant attentional processing independent of ADHD symptoms. *Journal of Neural Transmission*, *113*, 685–693.

Chen, W.J., Faraone, S.V., Biederman, J., & Tsuang, M.T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology*, *62*, 1017–1025.

Coghill, D., & Sonuga-Barke, E.J.S. (2012). Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders: implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, *53*, 469–489.

Curran, P.J., & Bauer, D.J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

Derks, E.M., Hudziak, J.J., Dolan, C.V., van Beijsterveldt, T.C.E.M., Verhulst, F.C., & Boomsma, D.I. (2008). Genetic and environmental influences on the relation between attention problems and attention deficit hyperactivity disorder. *Behavior Genetics*, *38*, 11–23.

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). New York: Oxford University Press.

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). Cambridge, MA: MIT Press.

Dodge, K.A., & Crick, N.R. (1990). Social information-processing bases of aggressive behavior in children. *Personality and Social Psychology Bulletin*, *16*, 8–22.

Duckworth, A.L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychological and Personality Science*, *1*, 311–317.

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 and Biobehavioral Reviews*, *25*, 355–373.

Gao, Y., & Raine, A. (2009). P3 event-related potential impairments in antisocial and psychopathic individuals: A meta-analysis. *Biological Psychology*, *82*, 199–210.

Gao, Y., Raine, A., Venables, P.H., & Mednick, S.A. (2013). The association between P3 amplitude at age 11 and criminal offending at age 23. *Journal of Clinical Child and Adolescent Psychology*, *42*, 120–130.

Gatzke-Kopp, L.M., Beauchaine, T.P., Shannon, K.E., Chipman, J., Fleming, A.P., Crowell, S.E., ... & Aylward, E. (2009). Neurological correlates of reward responding in adolescents with and without externalizing behavior disorders. *Journal of Abnormal Psychology*, *118*, 203–213.

Hermens, D.F., Ward, P.B., Hodge, M.A.R., Kaur, M., Naismith, S.L., & Hickie, I.B. (2010). Impaired MMN/P3a complex in first-episode psychosis: Cognitive and psychosocial associations. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*, 822–829.

Hicks, B.M., Bernat, E., Malone, S.M., Iacono, W.G., Patrick, C., Krueger, R.F., & McGue, M. (2007). Genes mediate the association between P3 amplitude and externalizing disorders. *Psychophysiology*, *44*, 98–105.

Hollingshead, A.B. (1975). Four factor index of social status: Yale University, Department of Sociology.

Hoyniak, C.P., Bates, J.E., Petersen, I.T., Yang, C.-L., Darcy, I., & Fontaine, N.M.G. (2018). Reduced neural responses to vocal fear: A potential biomarker for callous-uncaring traits in early childhood. *Developmental Science*, *21*, e12608.

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*, 155–166.

Iacono, W.G., & Malone, S.M. (2011). Developmental endophenotypes: Indexing genetic risk for substance abuse with the P300 brain event-related potential. *Child Development Perspectives*, *5*, 239–247.

Johnstone, S.J., Barry, R.J., Anderson, J.W., & Coyle, S.F. (1996). Age-related changes in child and adolescent event-related potential component morphology, amplitude and latency to standard and target stimuli in an auditory oddball task. *International Journal of Psychophysiology*, *24*, 223–238.

Johnstone, S.J., Barry, R.J., & Clarke, A.R. (2013). Ten years on: A follow-up review of ERP research in attention-deficit/hyperactivity disorder. *Clinical Neurophysiology*, *124*, 644–657.

Lifford, K.J., Harold, G.T., & Thapar, A. (2008). Parent-child relationships and ADHD symptoms: A longitudinal analysis. *Journal of Abnormal Child Psychology*, *36*, 285–296.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, *86*, 852–875.

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

Patrick, C.J. (2008). Psychophysiological correlates of aggression and violence: An integrative review. *Philosophical Transactions: Biological Sciences*, *363*, 2543–2555.

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*, 791–818.
Pogarell, O., Padberg, F., Karch, S., Segmiller, F., Juckel, G., Mulert, C., … & Koch, W. (2011). Dopaminergic mechanisms of target detection — P300 event related potential and striatal dopamine. *Psychiatry Research: Neuroimaging*, *194*, 212–218.
Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148.
Polich, J. (2012). Neuropsychology of P300. In S.J. Luck & E.S. Kappenman (Eds.), *Oxford handbook of event-related potential components* (pp. 159–188). New York: Oxford University Press.
Rydkjær, J., Møllegaard Jepsen, J.R., Pagsberg, A.K., Fagerlund, B., Glenthøj, B.Y., & Oranje, B. (2017). Mismatch negativity and P3a amplitude in young adolescents with first-episode psychosis: A comparison with ADHD. *Psychological Medicine*, *47*, 377–388.
Sattler, J.M., & Hoge, R.D. (2006). *Assessment of children: Behavioral, social, and clinical foundations* (5th edn). San Diego, CA: Jerome M. Sattler.
Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
Soltani, M., & Knight, R.T. (2000). Neural origins of the P300. *Critical Reviews in Neurobiology*, *14*, 199–224.
Tremblay, R.E. (2002). Prevention of injury by early socialization of aggressive behavior. *Injury Prevention*, *8*, iv17–iv21.
Tsai, M.-L., Hung, K.-L., & Lu, H.-H. (2012). Auditory event-related potentials in children with attention deficit hyperactivity disorder. *Pediatrics and Neonatology*, *53*, 118–124.
Turetsky, B.I., Dress, E.M., Braff, D.L., Calkins, M.E., Green, M.F., Greenwood, T.A., … & Light, G. (2015). The utility of P300 as a schizophrenia endophenotype and predictive biomarker: Clinical and socio-demographic modulators in COGS-2. *Schizophrenia Research*, *163*, 53–62.
van Beijsterveldt, C.E.M., & van Baal, G.C.M. (2002). Twin and family studies of the human electroencephalogram: A review and a meta-analysis. *Biological Psychology*, *61*, 111–138.
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*, e87347.
Yoon, H.H., Iacono, W.G., Malone, S.M., Bernat, E.M., & McGue, M. (2008). The effects of childhood disruptive disorder comorbidity on P3 event-related brain potentials in preadolescents with ADHD. *Biological Psychology*, *79*, 329–336.
Yoon, H.H., Malone, S.M., & Iacono, W.G. (2015). Longitudinal stability and predictive utility of the visual P3 response in adults with externalizing psychopathology. *Psychophysiology*, *52*, 1632–1645.

Accepted for publication: 17 August 2018

Supplementary Appendix S1. Missingness.

Some data were missing because of planned missingness ($n_{30\ months}$ = 17, $n_{36\ months}$ = 7) and some data were missing because of censoring (i.e., a child had not yet aged into a given measurement occasion, and therefore was not yet eligible; $n_{36\ months}$ = 9; $n_{42\ months}$ = 36). Censoring and planned missingness reflect data that are missing completely at random (MCAR). Planned missingness is an a priori design feature in which not all participants will be invited to participate at all time points. A portion of the children participated in a prior version of the protocol at the earliest ages (i.e., 30 and 36 months of age) while participating in the current protocol at later ages (i.e., 36 and 42 months of age). Using a planned missingness design, we opted beforehand to include these children's data in any longitudinal analyses (including the present study) even though they had not completed the current version of the protocol at the earliest ages. However, our hierarchical linear modeling strategy handles missing data, and our findings were substantially similar when using multiple imputation approaches or when excluding participants who had a P3 amplitude at only one time point (see Supplementary Appendix S3), suggesting that the missingness did not substantially influence our inferences.

Of the children with usable EEG data, 25 had usable EEG data at all three ages, and 54 children provided usable data at only two time points. Including the 25 children who provided usable EEG data at all three ages, 50 total children provided usable EEG data at 30 and 36 months, 39 children at 36 and 42 months, and 40 children at 30 and 42 months. Excluding the children who provided usable EEG data at all three ages, 25 children provided usable EEG data at only 30 and 36 months, 14 children at only 36 and 42 months, and 15 children at only 30 and 42 months. These numbers are displayed in Supplementary Table S1. Reasons for missingness in EEG data (i.e., a child did not provide usable data despite a scheduled EEG visit) were as

follows: child refusal to wear the cap (4% of the sample), child refusal to participate in the task (2%), fewer than 100 electrodes with low impedances (1%), fewer than 8 artifact-free trials in either condition of the oddball ERP task (3%), and other technical problems (4%). These attrition rates are consistent with other studies using EEG in young children (Bell & Cuevas, 2012).

We examined whether EEG missingness was related to other variables. There was no significant difference between children who did and did not provide usable EEG data in terms of the socioeconomic status (SES) of the child's family, or the child's age, sex, or level of temperamental inhibitory control or fear on the Children's Behavior Questionnaire (Rothbart, Ahadi, & Hershey, 1994). Missingness in externalizing problems was also unrelated to these variables, with one exception: missingness was higher in older children than younger children ($t[40.13] = -5.37$, $p < .001$), likely owing to sample attrition and censoring (discussed above).

Supplementary Appendix S2. Temporo-spatial principal components analysis.

Sequential temporo-spatial principal components analysis (PCA) is more likely to isolate the underlying neural signal when compared to examining individual electrodes (Dien & Frishkoff, 2005), and is especially useful when analyzing data from young children because of movement artifacts (Dien, 2012). Sequential temporo-spatial PCA was conducted using the ERP PCA toolkit (Dien, 2010). The number of factors to retain was selected objectively using a Parallel Test, which compares eigenvalues from the observed data to eigenvalues from randomly simulated data (Dien & Frishkoff, 2005). The PCA resulted in considerable data reduction from the original 38,400 data points per subject based on 128 electrodes × 300 time samples (i.e., 250 Hz sampling: 1 time sample every 4 ms of the 1,200 ms epoch = 1,200 ÷ 4). The initial temporal PCA identified 18 temporal factors accounting for 95% of variance from the 300 time samples, and the subsequent temporo-spatial PCA identified 13 spatial factors accounting for 87% of variance across the 128 electrodes. Thus, the temporo-spatial PCA reduced the data from 38,400 data points per subject to 234 principal components (i.e., 18 temporal factors × 13 spatial factors).

The temporo-spatial factor corresponding to the P3 (component 1 out of 234 components) was selected based on *a priori* expectations about the latency and topography of the component. Because this was the first temporo-spatial factor, it indicated that it explained the most variance in the waveform (8.35%). This temporo-spatial factor peaked around 448 ms post-stimulus, and included a posterior/parietal positivity with a corresponding frontal negativity. We focused on posterior/parietal electrodes (see Supplementary Figure S1) to identify this temporo-spatial factor because previous findings indicate that the P3 elicited from oddball paradigms has a posterior/parietal distribution in children (Hoyniak, Petersen, McQuillan, Staples, & Bates, 2015;

Johnstone, Barry, Anderson, & Coyle, 1996). PCA-derived P3 waveforms are depicted in Supplementary Figure S1. The PCA successfully isolated the P3 component (that distinguished between frequent and infrequent trials) from surrounding ERP components. Children's P3 amplitudes were calculated as their mean amplitudes for this temporo-spatial component. No windowing was used in the calculation; in PCA, *every* electrode and time sample is included in the calculation of the amplitudes *to the extent that* each electrode and time sample reflects the component of interest (in this case, component 1 reflecting the P3) based on factor loadings.

Supplementary Appendix S3. Information about the HLM growth curve models.

Model Equations for Final Model

Nesting:
Level 1: t = time (i.e., age in months, centered at 30 months of age; T1 = 0, T2 = 6, T3 = 12)
Level 2: i = individual

Level 1—Within individual:
$EXT_{ti} = \beta_{0i} + \beta_{1i}(time_{ti}) + \beta_{2i}(P3centered_{ti}) + \beta_{3i}(BadChannels_{ti}) + \beta_{4i}(InfrequentTrials_{ti}) + e_{ti}$

Level 2—Between individual:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}(sex_i) + \gamma_{02}(P3mean_i) + r_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}(sex_i) + \gamma_{12}(P3mean_i) + r_{1i}$$
$$\beta_{2i} = \gamma_{20}$$
$$\beta_{3i} = \gamma_{30}$$
$$\beta_{4i} = \gamma_{40}$$

Random effects (Level 1 variance component):
σ^2 = Level 1 variance

Random effects (Level 2 variance-covariance components):
τ_{00} = variance of the intercepts
τ_{01} = covariance between the intercepts and slopes
τ_{11} = variance of the slopes

Definitions of Model Parameters

"EXT" = externalizing problems; "P3centered" = child's P3 amplitude at a given time point that is centered around their own mean P3 amplitude across time (i.e., person-mean-centered P3 amplitude); "BadChannels" = number of bad channels; "InfrequentTrials" = number of infrequent trials kept; "P3mean" = a given child's mean P3 amplitude across time.

HLM growth curve models fit random intercepts and slopes of externalizing problems. "β_{0i}" reflects the association with the random intercepts, centered at 30 months (i.e., the first time point). "β_{1i}" reflects the association with the random linear slopes. "β_{2i}", "β_{3i}", and "β_{4i}" reflect the time-varying predictors. "γ_{00}" reflects the average intercept across children. "γ_{10}" reflects the average linear slope across children. "e_{ti}" reflects the Level-1 random effect (Level-

1 residuals). "r_{0i}" and "r_{1i}" reflect the Level-2 random effects (Level-2 residuals) of intercepts and slopes, respectively.

To disaggregate within- and between-individual effects in the association between the P3 and externalizing problems, we included both within-individual effects and between-individual effects in the HLM growth curve models, consistent with best practices (Curran & Bauer, 2011). The between-individual effects were estimated by including the child's own mean P3 amplitude across time ($P3mean_i$) as a predictor of both the intercepts (γ_{02}) and slopes (γ_{12}) of externalizing problems (i.e., Level 2). Both between-individual effects were estimated to disentangle the between-individual effects on the level (intercepts) versus change (slopes) of externalizing problems. The within-individual effect was estimated by centering the P3 amplitude around the child's own mean, so-called person-mean centering. That is, a child's own mean P3 amplitude across time ($P3mean_i$) was subtracted from their own time-specific P3 amplitude: i.e., $P3centered_{ti} = P3_{ti} - P3mean_i$. The within-individual effect was estimated by including the person-mean-centered P3 amplitude ($P3centered_{ti}$) as a time-varying predictor (γ_{20}) of externalizing problems (i.e., Level 1).

We were primarily interested in examining both the within-individual (γ_{20}) and between-individual (γ_{02}, γ_{12}) associations between the P3 and externalizing problems. "γ_{20}" reflects whether within-individual changes in the P3 amplitude (i.e., individuals' time-specific deviations in the P3 amplitude away from their own mean) are associated with concomitant within-individual changes in externalizing problems (time-specific deviations in externalizing problems over and above their linear slopes, β_{1i}). "γ_{02}" reflects whether a child's mean P3 amplitude across time is associated with their intercepts of externalizing problems (i.e., their level at 30

months). "γ_{12}" reflects whether a child's mean P3 amplitude across time is associated with their linear slopes of externalizing problems across 30 to 42 months.

We examined within- and between-individual associations in the context of covariates that may account for any potentially systematic ERP missingness (the number of infrequent trials kept and the number of bad channels). We also included the child's sex as a covariate given the sex differences in rates and growth of aggression and ADHD symptoms in early childhood (Baillargeon et al., 2007). We allowed all time-invariant predictors to covary with both the intercepts and slopes of externalizing problems, so we could comparably compare the between-child association of the child's mean P3 amplitude across time with their intercepts versus slopes of externalizing problems. In terms of time-invariant covariates, "γ_{01}" reflects whether boys and girls differ in their intercepts of externalizing problems (i.e., their level at 30 months). "γ_{11}" reflects whether boys and girls differ in their linear slopes of externalizing problems across 30 to 42 months. In terms of time-varying covariates, "γ_{30}" reflects whether the number of bad channels is associated with externalizing problems, and "γ_{40}" reflects whether the number of infrequent trials kept is associated with externalizing problems.

Additional Information about the HLM Growth Curve Models

We fit growth curve models using the lme function of the nlme package (Pinheiro, Bates, DebRoy, & Sarkar, 2009) in R (R Core Team, 2016) for hierarchical linear modeling (HLM). The HLM growth curve models in the present study fit random intercepts and slopes of externalizing problems. The study aimed to estimate both within-individual and between-individual associations between the P3 and externalizing problems. To disaggregate within- and between-individual effects in the models, we included predictors for both the child's mean P3 amplitude across time (i.e., a between-individual effect) and the child's person-mean-centered P3

amplitude (i.e., P3 amplitude at a given time point that is centered around their own mean P3 amplitude across time; a within-individual effect). Thus, the HLM growth curve models examine both the within-individual and between-individual association between the predictor (P3) and outcome (externalizing problems). At the between-individual level, the HLM growth curve models examine children's mean values on the predictor across time in relation to their intercept and slopes on the outcome over time. At the within-individual-level, the HLM growth curve models examine individuals' time-specific deviations in the predictor in relation to their time-specific deviations in the outcome *over and above* the effect of time (i.e., β_{1i}, linear slopes) on the outcome (Curran, Lee, Howard, Lane, & MacCallum, 2012; Wang & Maxwell, 2015). Thus, our model examines whether within-individual changes in the predictor (P3 amplitude) are associated with concomitant within-individual changes in the outcome (externalizing problems; Duckworth, Tsukayama, & May, 2010; Galla et al., 2014). The time-varying P3 predictor (P3centered_{ti}) reflects children's change (i.e., time specific deviations) in P3 amplitude because the between-subject variance has been removed as it is person-mean centered, and therefore, children's changes in P3 amplitude relative to their own mean P3 amplitude are captured. The time-varying P3 predictor predicts children's concomitant change (i.e., time-specific deviations) in externalizing problems (γ_{20}) because the association is with a time-varying predictor whose between-subject variance has been removed, and therefore it only accounts for within-individual (i.e., Level-1) variance, i.e., change, in externalizing problems that is over and above the effect of their linear slopes of externalizing problems (β_{1i}). An assumption of this modeling approach is that the person-level mean P3 amplitude is estimated without error. For other empirical examples using similar models, see Duckworth et al. (2010) and Galla et al. (2014).

Model Building

To select final models, we followed a model building sequence. Consistent with recommendations (Hox, Moerbeek, & van de Schoot, 2017), restricted maximum likelihood was used to compare models with the same fixed effects but different random effects. Full maximum likelihood was used to compare models with different fixed effects. To account for the modest sample size, the final models used restricted maximum likelihood (McNeish & Stapleton, 2016).

First, we fit unconditional means models with random intercepts to the trajectories of externalizing problems. The unconditional means models of aggression showed considerable within-individual (σ^2 = 13.14; 95% confidence interval of standard error estimate: [3.30, 3.98]) and between-individual (τ_{00} = 20.42; 95% CI of standard deviation estimate: [3.89, 5.25]) variance, suggesting that the average child varies over time in aggression and that the cross-time means of aggression differ between children. We also observed significant within-individual (σ^2 = 1.31; 95% CI of standard error estimate: [1.04, 1.26]) and between-individual (τ_{00} = 1.60; 95% CI of standard deviation estimate: [1.08, 1.49]) variance in ADHD symptoms, suggesting that the average child varies over time in ADHD symptoms and that the cross-time means of ADHD symptoms differ between children. However, there was less within- and between-individual variance in ADHD symptoms compared to aggression, possibly in part because the scale of ADHD symptoms was shorter than the aggression scale.

Second, to account for the change in externalizing problems over time, an unconditional growth model was fit with random intercepts and a linear random slope for time (a random intercepts and slopes model). Despite a non-significant mean of the slopes for aggression or ADHD, there was a significant variance of the slopes for both aggression (95% CI of standard deviation estimates: τ_{11} = [0.11, 0.44]) and ADHD symptoms (95% CI: τ_{11} = [0.08, 0.15]), indicating that children showed significant within-child change in aggression and ADHD

symptoms over time (i.e., children showed different trajectories). Although models with random slopes did not fit significantly better than models with fixed slopes for aggression at the $p < .05$ level ($\chi^2[2] = 4.34$, $p = .11$), we retained random slopes in order to (a) most accurately estimate each child's growth curve, and (b) fit models consistent with best practices for disaggregating within- and between-individual associations. As Curran and Bauer (2011, p. 589) note, "An important element of the growth model is that the values of the intercept and slope components vary randomly across persons." Models with random slopes fit significantly better than models with fixed slopes for ADHD symptoms ($\chi^2[2] = 13.40$, $p = .001$). Findings did not substantially differ in the final models when fitting fixed slopes rather than random slopes.

Third, we added predictors for the within-individual ($P3centered_{ti}$) and between-individual ($P3mean_i$) associations between P3 amplitudes and externalizing problems. Fourth, we added covariates that might plausibly explain the association between P3 amplitudes and externalizing problems (sex_i, $BadChannels_{ti}$, and $InfrequentTrials_{ti}$). The growth curve models with predictors for the within-individual and between-individual associations between P3 amplitudes and externalizing problems, along with covariates, served as the final models.

Multiple Imputation

As a sensitivity analysis, we examined models with multiple imputation using the Amelia package (Honaker, King, & Blackwell, 2011) in R. Amelia uses an expectation-maximization with bootstrapping algorithm, and is well suited for longitudinal data (Honaker & King, 2010). All model variables—age, sex, oddball P3 amplitude, aggression, covariates for ADHD symptoms, numbers of good trials and bad channels—were used to create imputed values for 1,000 data sets. The variables identifying the participant and participant's age were specified in the imputation model to appropriately handle the dependence of longitudinal data. Data were

imputed 1,000 times to ensure appropriate power (Graham, Olchowski, & Gilreath, 2007), but findings did not substantially differ with fewer imputations. The HLM growth curve models were run on each imputed data set separately, and then the results were combined using the mitools (Lumley, 2010) and mix (Schafer, 1997) packages in R, which use Rubin's (1987) rules for combining results of analyses on multiply imputed data sets. Because the focal results were substantially unchanged when using multiple imputation (although the between-individual association between the child's mean P3 amplitude across time and their intercepts of aggression became marginally significant, the within-individual association between person-mean-centered P3 amplitudes and aggression remained significant; see Supplementary Table S4), results from the raw data are presented.

Follow-Up Analyses

As a follow-up test, we also examined the higher-order Externalizing Problems scale, composed of the sum of the Aggression and Attention Problems first-order scales. Within-child changes in the P3 amplitude were negatively associated with within-child changes in externalizing problems (γ_{20}: B = -.36, β = -.11, SE = .13, df = 90, p = .006). Thus, findings for the Externalizing Problems scale closely paralleled results for the Aggression scale.

As another follow-up analysis, we examined whether the within-individual association between P3 amplitudes and aggression differed between boys and girls. The association did not significantly differ between boys and girls (γ_{21}: B = .08, β = .02, SE = .22, df = 89, p = .699). Nevertheless, we likely have limited power to detect sex differences in the association, so we interpret these results with caution.

To examine the extent to which missingness impacts our results, we fit separate models with only those children who had a P3 amplitude at two or more time points (children who

inform the estimate of the time-varying effect, i.e., the within-child association), in addition to models with multiple imputation (see above). Model results were the same when we fit the models only with children who had a P3 amplitude at two or more time points. Thus, excluding participants who did not inform the time-varying effect did not impact our findings.

Model Assumptions

One of the assumptions of disaggregating within- and between-individual effects using person-mean centering in a HLM growth curve model is that the time-varying predictor does not show systematic differences across time (Curran & Bauer, 2011). To test this assumption, we examined whether P3 amplitudes significantly differed across time using paired sample *t*-tests. P3 amplitudes did not significantly differ from 30 to 36 ($t[49] = 1.38$, $p = .175$) or from 36 to 42 ($t[38] = -0.27$, $p = .785$) months. Thus, to disaggregate within- and between-child effects in the association of P3 amplitudes with externalizing problems, we met the assumption that P3 amplitudes did not show significant differences across time. Nevertheless, even if an undetected trend of the P3 amplitude exists, we followed best practices for handling this possibility by simultaneously examining (a) the child's mean P3 amplitude (γ_{02}, γ_{12}), (b) the child's time-varying person-mean centered P3 amplitude (γ_{20}), and (c) time (γ_{10}) as a time-varying predictor (Wang & Maxwell, 2015). In sum, evidence that (a) we obtained similar findings with and without multiple imputation, and that (b) we met model assumptions, provides greater confidence in our inferences.

Supplementary Table S1. EEG observations per child by number of assessments and age at time of visit.

	Age at Assessment			
Usable EEG data	30	36	42	Number of Children
1 assessment (*n* = 74)	41			41
		14		14
			19	19
2 assessments (*n* = 54)	25	25		25
		14	14	14
	15		15	15
3 assessments (*n* = 25)	25	25	25	25
Usable EEG observations per age at assessment	106	78	73	153

Supplementary Table S2. Descriptive statistics and correlation matrix of P3 amplitude, aggression, and ADHD symptoms at 30, 36, and 42 months of age.

	P3 Amplitude (μV)			Aggression			ADHD symptoms		
	30 mo	36 mo	42 mo	30 mo	36 mo	42 mo	30 mo	36 mo	42 mo
P3 – 30 mo	–								
P3 – 36 mo	.15	–							
P3 – 42 mo	.38*	.33*	–						
AGG – 30 mo	-.14	.02	-.19	–					
AGG – 36 mo	.11	-.17	-.13	.55***	–				
AGG – 42 mo	.09	-.23†	-.18	.52***	.75***	–			
ADHD – 30 mo	-.04	-.01	-.10	.50***	.31***	.25*	–		
ADHD – 36 mo	.16	-.11	-.12	.37***	.60***	.46***	.58***	–	
ADHD – 42 mo	.11	-.09	-.11	.23*	.36***	.51***	.39***	.63***	–
M	5.13	4.32	4.32	9.14	10.20	9.21	2.51	2.36	2.10
SD	3.80	3.98	3.85	5.55	5.95	5.82	1.82	1.55	1.63

Note. *** $p < .001$; * $p < .05$; † $p < .10$; all *p*s two-tailed. "mo" = months of age. It is interesting that there is no simple bivariate association between P3 amplitudes and either aggression or ADHD at any age, even though the associations have the same direction (negative) and magnitude (*r*s between -.14 and -.18 for aggression) as the direction and magnitude of the association that collapses across age and was significantly negative ($r = -.16$, $p = .011$; Supplementary Table S3). The bivariate associations also have the same direction and magnitude as the direction and magnitude of the significant association observed in a HLM model that accounts for nested data ($\beta = -.12$, $p = .004$; Table 2). This discrepancy may reflect the attenuated power at any given age (because of a smaller *n*). This would emphasize the importance of longitudinal analytic approaches that use all available information for increased power along with reduced measurement error (Scherbaum & Ferreter, 2009), such as the longitudinal analytic approaches used in the present study.

Supplementary Table S3. Descriptive statistics and correlation matrix of model variables.

Variable	Age	Sex	P3	Aggression	ADHD	Number of Bad Channels	Infrequent Trials Kept
Age	–						
Sex	.00	–					
P3 ERP amplitude	-.09	-.12†	–				
Aggression	.01	-.09†	-.16*	–			
ADHD symptoms	-.10†	-.11*	-.07	.53***	–		
Number of Bad Channels	.00	-.23***	.00	.07	.10	–	
Infrequent Trials Kept	-.13*	.09	.01	-.10	-.07	-.33***	–
M	36.00	0.45	4.66	9.50	2.35	10.07	25.30
SD	4.90	0.50	3.87	5.76	1.69	3.83	7.67

Note. *** $p < .001$; * $p < .05$; † $p < .10$; all *p*s two-tailed. "Sex" is coded with female = 1, male = 0. Child age in months. A data dictionary of the study variables is published at https://osf.io/25vq6.

Supplementary Table S4. Results of HLM growth curve models with multiple imputation.

Outcome: Aggression	*B*	*β*	*SE*	*df*	*p*
intercept	**13.55**	**0.03**	**2.38**	**3,623.54**	**< .001**
Time	-0.02	0.00	0.14	8,733.04	.876
Sex	-1.38	-0.12	0.94	64,582.89	.144
Sex × Time	0.00	0.00	0.10	14,164.79	.976
Mean P3 amplitude	-0.37	-0.15	0.22	9,209.97	.096
Mean P3 amplitude × Time	0.00	0.00	0.02	8,359.31	.987
***Time-varying P3 amplitude**	**-0.20**	**-0.10**	**0.09**	**3,910.25**	**.039**
*Number of bad channels	-0.01	0.00	0.10	3,160.58	.931
*Number of infrequent trials kept	-0.06	-0.08	0.05	2,706.79	.236

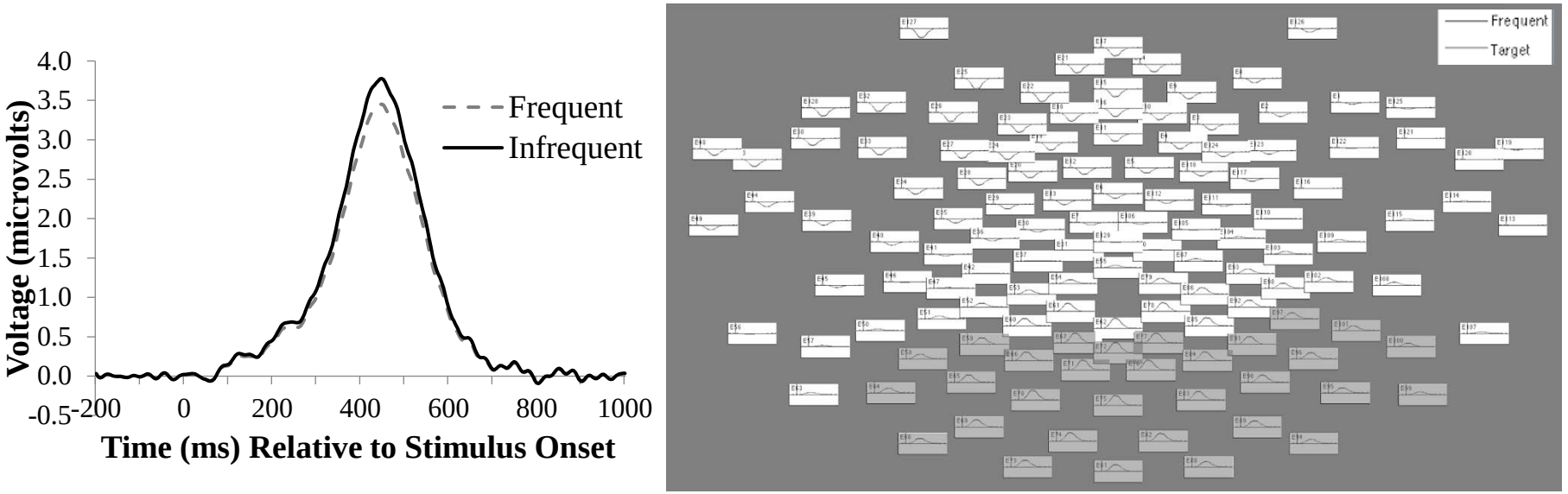
Outcome: ADHD symptoms	*B*	*β*	*SE*	*df*	*p*
intercept	**3.12**	**0.00**	**0.69**	**5,961.24**	**< .001**
Time	-0.07	-0.11	0.04	12,038.20	.116
Sex	**-0.62**	**-0.12**	**0.30**	**122,430.17**	**.039**
Sex × Time	0.04	0.05	0.03	22,492.27	.283
Mean P3 amplitude	-0.07	-0.07	0.07	11,950.96	.346
Mean P3 amplitude × Time	0.00	0.02	0.01	10,475.15	.705
*Time-varying P3 amplitude	-0.03	-0.06	0.03	5,124.90	.216
*Number of bad channels	0.02	0.03	0.03	4,924.21	.584
*Number of infrequent trials kept	-0.01	-0.02	0.01	3,891.39	.705

Note: "Time" reflects the slope term, and is centered at the first time point so that the intercept reflects the child's level at 30 months (i.e., 0, 6, 12 months from 30 months). "Sex" is coded with female = 1, male = 0. "Mean P3 amplitude" refers to a given child's mean P3 amplitude across time (time invariant). "Time-varying P3 amplitude" refers to a given child's P3 amplitude at a given time point that is centered around their mean P3 amplitude across time (time varying). Results were combined from 1,000 multiply imputed data sets (see Supplementary Appendix S3). Interaction terms with time essentially reflect the association with the slopes of the outcome (e.g., "Sex × Time" reflects sex in association with the slopes of the outcome). Asterisks reflect time-varying terms. Terms in bold reflect significant associations at $p < .05$ level.

Supplementary Table S5. Results of HLM growth curve model of aggression while accounting for ADHD symptoms (and additional covariates).

Outcome: Aggression	*B*	*β*	*SE*	*df*	*p*
intercept	**9.67**	**-0.01**	**2.02**	**142**	**< .001**
Time	0.00	0.02	0.11	89	.971
Sex	0.07	-0.03	0.86	142	.932
Sex × Time	-0.08	-0.04	0.11	89	.441
Mean P3 amplitude	**-0.25**	**-0.10**	**0.12**	**142**	**.045**
Mean P3 amplitude × Time	0.02	0.04	0.02	89	.365
***Time-varying P3 amplitude**	**-0.25**	**-0.09**	**0.09**	**89**	**.007**
***ADHD symptoms**	**1.66**	**0.51**	**0.17**	**89**	**< .001**
*Number of bad channels	-0.07	-0.04	0.07	89	.380
***Number of infrequent trials kept**	**-0.10**	**-0.11**	**0.05**	**89**	**.032**

Note: We examined growth curve models of aggression while accounting for ADHD symptoms to determine whether the association between P3 amplitudes and aggression was due to the comorbidity of aggression with attention problems, to advance our understanding of the specificity of the effects of the P3. "Time" reflects the slope term, and is centered at the first time point so that the intercept reflects the child's level at 30 months (i.e., 0, 6, 12 months from 30 months). "Sex" is coded with female = 1, male = 0. "Mean P3 amplitude" refers to a given child's mean P3 amplitude across time (time invariant). "Time-varying P3 amplitude" refers to a given child's P3 amplitude at a given time point that is centered around his or her mean P3 amplitude across time (time varying). Interaction terms with time essentially reflect the association with the slopes of the outcome (e.g., "Sex × Time" reflects sex in association with the slopes of the outcome). Asterisks reflect time-varying terms. Terms in bold reflect significant associations at $p < .05$ level.



Supplementary Figure S1. Left panel: Children's PCA-derived P3 waveforms. The waveform depicted represents the mean P3 waveform from those electrodes with a 0.6 or greater factor loading onto the PCA component reflecting the P3 (i.e., the first tempor-spatial PCA component; see Supplementary Appendix S2). For purposes of depicting waveforms, electrodes from the PCA-derived posterior/parietal electrode cluster (see right panel with gray-shaded electrodes) were averaged with equal, unit weighting. However, actual P3 amplitudes were calculated using PCA, in which all electrodes contribute to estimation of amplitudes to the extent that they reflect the underlying P3 component (based on factor loadings), thus accentuating those electrodes driving the signal. This accounts for larger amplitudes in Supplementary Tables S2 and S3 than Figure 1 and Supplementary Figure S1.

Right panel: Topo plot of the PCA component reflecting the P3. Waveforms from electrodes shaded in gray have a 0.6 or greater factor loading onto the PCA component reflecting the P3.

References

Baillargeon, R. H., Normand, C. L., Séguin, J. R., Zoccolillo, M., Japel, C., Pérusse, D., . . . Tremblay, R. E. (2007). The evolution of problem and social competence behaviors during toddlerhood: A prospective population-based cohort survey. *Infant Mental Health Journal, 28*, 12-38. doi: doi:10.1002/imhj.20120

Bell, M. A., & Cuevas, K. (2012). Using EEG to study cognitive development: Issues and practices. *Journal of Cognition and Development, 13*, 281-294. doi: 10.1080/15248372.2012.691143

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62*, 583-619. doi: 10.1146/annurev.psych.093008.100356

Curran, P. J., Lee, T., Howard, A. L., Lane, S., & MacCallum, R. C. (2012). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. In G. R. Hancock & J. R. Harring (Eds.), *Advances in longitudinal methods in the social and behavioral sciences*.

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,* 138-145. doi: 10.1016/j.jneumeth.2009.12.009

Dien, J. (2012). Applying principal components analysis to event-related potentials: A tutorial. *Developmental Neuropsychology, 37*, 497-517. doi: 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). Cambridge, MA, US: MIT Press.

Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychological and Personality Science, 1*, 311-317. doi: 10.1177/1948550609359707

Galla, B. M., Wood, J. J., Tsukayama, E., Har, K., Chiu, A. W., & Langer, D. A. (2014). A longitudinal multilevel model analysis of the within-person and between-person effect of effortful engagement and academic self-efficacy on academic performance. *Journal of School Psychology, 52*, 295-308. doi: 10.1016/j.jsp.2014.04.001

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

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science, 54*, 561-581. doi: 10.1111/j.1540-5907.2010.00447.x

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45*, 1-47. doi: 10.18637/jss.v045.i07

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.): Taylor & Francis.

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*, 155-166. doi: 10.1080/87565641.2015.1016162

Johnstone, S. J., Barry, R. J., Anderson, J. W., & Coyle, S. F. (1996). Age-related changes in child and adolescent event-related potential component morphology, amplitude and latency to standard and target stimuli in an auditory oddball task. *International Journal of Psychophysiology, 24*, 223-238. doi: 10.1016/s0167-8760(96)00065-7

Lumley, T. (2010). mitools: Tools for multiple imputation of missing data. R package version 2. Retrieved from http://cran.r-project.org/web/packages/mitools/

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*, 295-314. doi: 10.1007/s10648-014-9287-x

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D., and the R Core team. (2009). nlme: Linear and nonlinear mixed effects models. R package version 3.1-93. Retrieved from http://cran.r-project.org/web/packages/nlme/index.html

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. *Merrill-Palmer Quarterly, 40*, 21-39.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*: John Wiley & Sons, Inc.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (Vol. 72). New York, NY: Chapman & Hall/CRC.

Scherbaum, C. A., & Ferreter, J. M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods, 12*, 347-367. doi: 10.1177/1094428107308906

Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods, 20*, 63-83. doi: 10.1037/met0000030