Received: 15 June 2021 | Revised: 26 January 2022 | Accepted: 19 February 2022

DOI: 10.1111/psyp.14043

PSYCHOPHYSIOLOGY SPR WILEY

ORIGINAL ARTICLE

Reliability of reward ERPs in middle-late adolescents using a custom and a standardized preprocessing pipeline

György Hámori[1,2] | Alexandra Rádosi[1,3] | Bea Pászthy[4] | János M. Réthelyi[5] | István Ulbert[6,7] | Richárd Fiáth[6,7] | Nóra Bunford[1]

[1]Developmental and Translational Neuroscience Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary

[2]Department of Cognitive Science, Faculty of Natural Sciences, Budapest University of Technology and Economics, Budapest, Hungary

[3]Doctoral School of Mental Health Sciences, Semmelweis University, Budapest, Hungary

[4]1st Department of Paediatrics, Faculty of Medicine, Semmelweis University, Budapest, Hungary

[5]Department of Psychiatry and Psychotherapy, Faculty of Medicine, Semmelweis University, Budapest, Hungary

[6]Integrative Neuroscience Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Budapest, Hungary

[7]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary

Correspondence
Nóra Bunford, Developmental and Translational Neuroscience Research Group, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Magyar Tudósok körútja 2., H-1117 Budapest, Hungary.
Email: bunford.nora@ttk.hu

Funding information
Magyar Tudományos Akadémia, Grant/Award Number: LP2018-3/2018

Abstract

Despite advantage of neuroimaging measures in translational research frameworks, less is known about the psychometric properties thereof, especially in middle-late adolescents. Earlier, we examined evidence of convergent and incremental validity of reward anticipation and response event-related potentials (ERPs) and here we examined, in the same sample of 43 adolescents (M_{age} = 15.67 years; SD = 1.01; range: 14–18; 32.6% boys), data quality (signal-to-noise ratio [SNR]), stability (mean amplitude across trials), and internal consistency (Cronbach's α and split-half reliability) of the same ERPs. Further, because observed time course and peak amplitude of ERP grand averages and thus findings on SNR, stability, and internal consistency may depend on preprocessing method, we employed a custom and a standardized preprocessing pipeline and compared findings across those. Using our custom pipeline, reward anticipation components were stable by the 40th trial, achieved acceptable internal consistency by the 19th, and all (but the stimulus-preceding negativity [SPN]) achieved acceptable SNR by the 41st trial. Initial response to reward components were stable by the 20th trial and achieved acceptable internal consistency by the 11th and acceptable SNR by the 45th trial. Difference scores had worse psychometric properties than parent measures. Time course and peak amplitudes of ERPs and thus results on SNR, stability, and internal consistency were comparable across

Richárd Fiáth and Nóra Bunford contributed equally to this study.

preprocessing pipelines. In case of reward anticipation ERPs examined here, 41 trials (+4 artifacted and removed) and, in case of reward response ERPs, 45 trials (+5 artifacted) yielded stable and internally consistent estimates with acceptable SNR. Results are robust across preprocessing methods.

KEYWORDS

adolescent, ERP preprocessing pipeline, event-related potential, psychometric property, reward

1 | INTRODUCTION

Accumulating evidence supports the utility of neuroimaging[1] techniques in examining brain structure and function across typical and pathological populations (Gabrieli et al., 2015), and there has been an increase in application of such measures in experimental and clinical research. Amongst others, neuroimaging techniques have shown promise in the assessment and identification of individual differences (e.g., in learning) across development in typical populations (Gabrieli et al., 2015) and in predicting heterogeneity (Bunford, Kujawa, Swain, et al., 2017) and treatment response (Bunford, Kujawa, Fitzgerald, et al., 2017; Kujawa et al., 2016) in pediatric and adult psychiatric samples. What is more, neuroimaging indices have been shown to enhance, and, in some cases even outperform, traditional measures and predictors of treatment response (Gabrieli et al., 2015; Zubovics et al., 2021). Traditional measures, such as rating scales, educational or neuropsychological test scores, are often inconsistent in this regard (Layne et al., 2003) and traditional predictors, such as demographics, environmental factors, or symptoms (for review, see Bunford, Kujawa, Fitzgerald, et al., 2017), account for a relatively small portion of variance in treatment response. For example in adults, anxiety severity measured at pre-treatment accounts for 12% of the variance in cognitive behavioral therapy (CBT)-related anxiety change (Doehrmann et al., 2013; Whitfield-Gabrieli et al., 2015) whereas inclusion of neuroimaging indices close to doubles (Whitfield-Gabrieli et al., 2015) or triples (Doehrmann et al., 2013) the amount of variance accounted for.

Although these data underscore the potential value of neuroimaging measures in individual differences and clinical research, the exponential increase in the use of such measures has not been paralleled by inquiry into their psychometric properties (Siegle, 2011). Evidence of reliability and validity is key in experimental studies comparing conditions or groups (Thigpen et al., 2017) and is a precondition to clinical utility. For example, to be able to assess change (e.g., as a result of disease progression or treatment), a measure has to be stable over time (i.e., test–retest reliability) (Siegle, 2011). Similarly, to be able to assess association with (e.g., clinical) outcomes, data are needed on the amount of true score contained in a measure (as opposed to error variance) that is available for such association (i.e., internal consistency) (Levinson et al., 2017). Yet, neuroimaging measures have not been held to the same standards as traditional measures (Frewen et al., 2008), including with regard to basic psychometric properties such as test–retest reliability or internal consistency (Siegle, 2011), despite reliability being a prerequisite of validity (Cronbach & Meehl, 1955; Luking et al., 2017). As a result, the degree to which neuroimaging can directly impact clinical practice has been up for debate (Ball et al., 2014).

Our focus across our earlier (Zubovics et al., 2021) and current research is on psychometric properties of event-related potential (ERP) indices of reward processing in middle-late adolescents. ERPs are changes in the electroencephalogram (EEG) linked to specific events (e.g., presentation of a stimulus), reflecting synchronous activity of neuronal populations (Hajcak et al., 2011). Reward processing is of interest as differences in reward system regulation predicts development of externalizing and internalizing disorders (Bunford et al., 2021; Kujawa et al., 2018) as well as increases in depressive symptoms and substance abuse (Bress et al., 2013; Morgan et al., 2013; Stice et al., 2013). Middle-late adolescence is an especially relevant developmental phase from the perspective of reward processing as adolescent neuromaturation is such that there is a developmental discrepancy between brain regions implicated in the generation and the regulation of appetitive behavior (Ernst & Spear, 2009; Galván, 2013; Kringelbach, 2005; Spear, 2013, 2018), contributing to within-person peak in reward sensitivity during adolescence and a between-person increase in reward but decrease in punishment sensitivity (in adolescents relative to children and adults) (Cauffman et al., 2010; Ernst, 2014; Ernst & Spear, 2009; Shulman et al., 2016). These horizontal and vertical differences make adolescence a

[1]In keeping with the literature, we use the term "neuroimaging" as inclusive of the event-related potential (ERP) technique. We nevertheless recognize that this is not entirely accurate, as the ERP technique does not yield an actual image of the brain (Luck, 2014).

sensitive window into the effects of atypical reward processing (Bress et al., 2013; Bunford et al., 2021; Kujawa et al., 2018; Morgan et al., 2013; Silverman et al., 2015; Stice et al., 2013).

Regarding psychometric properties of ERPs in general, there has been increased attention devoted to assessing the psychometric properties of ERPs and findings are encouraging; e.g., evidence shows ERPs are stable across time (e.g., 3-month retest in Auerbach et al., 2016; 1-month retest in Cassidy et al., 2012; 20-min retest in Segalowitz et al., 2010), internally consistent (e.g., Marco-Pallares et al., 2011) and correspond to relevant characteristics and pathologies as expected (e.g., Klawohn et al., 2020; Kujawa et al., 2019; Kujawa, Proudfit, Kessel, et al., 2015). Regarding psychometric properties of ERPs of reward processing, despite evidence that characteristics of reward ERPs differ across development (both the reward positivity [RewP] and the feedback negativity [FN] attenuate across the life span (Hämmerer et al., 2011) and both the amplitude (Hämmerer et al., 2011; Lukie et al., 2014) and the latency (Crowley et al., 2013) of the FN change with development), the majority of the available literature is based on adult research and there is a relative paucity of empirical work assessing the psychometric properties of reward ERPs in middle-late adolescents.

Next, we review the available literature assessing the psychometric properties of reward ERPs across age groups; instead of aiming for a comprehensive review our goal is to highlight areas where relevant data are available (with illustrative examples) and areas where a paucity of research is apparent. We focus on two aspects of reward processing, reward anticipation and initial response to reward attainment (hereafter: reward response) as probed by the monetary incentive delay (MID) (Knutson et al., 2003; Knutson, Fong, Adams, et al., 2001; Knutson, Fong, Hommer, et al., 2001) and the Doors (Dunning & Hajcak, 2007; Kujawa et al., 2018; Kujawa, Proudfit, Hajcak, et al., 2015) tasks, respectively. Regarding MID ERPs, the Cue P3 measures attention allocation to cue, modulated by affective significance and the reward value of stimuli (Chronaki et al., 2017). The Target P3, comparably the Cue P3, reflects motivated, task-relevant attention to a target as well as stimulus categorization and evaluation (Broyd et al., 2012; Groom et al., 2010). The stimulus preceding negativity (SPN), a slowly growing negativity that reaches its maximum prior to stimulus onset, measures anticipatory processes, including anticipatory attention or anticipation of the affective valence of a(n informative) feedback (Foti & Hajcak, 2012). In both the MID and the Doors task, in line with others (Luking et al., 2017), we examined the RewP, a relative positivity following gains; the FN, a relative negativity following losses; and the ΔRewP, a win-loss difference score that manifests as a relative positivity in the ERP waveform following feedback associated with neural activity linked to initial reward response (Foti et al., 2011; Kujawa et al., 2014; Kujawa et al., 2018). Of note, there is some debate about appropriateness of the MID task to probe both reward anticipation and response (National Institute of Mental Health et al., 2016), despite empirical findings indicating it is ideal for differentiating reward anticipation from response (Knutson, Fong, Adams, et al., 2001; Novak et al., 2016; Novak & Foti, 2015). We interpret ERPs preceding reward receipt (Cue P3, Target P3, SPN) as reflecting anticipation and ERPs following reward/loss (RewP, FN, ΔRewP) as reflecting initial response to reward outcome.

1.1 | Psychometric properties of MID and Doors ERPs

1.1.1 | Reliability

We are aware of no studies evaluating evidence of reliability of MID ERPs in children and one study evaluated evidence of internal consistency of MID ERPs in adolescents and one in adults. In adolescents, Cue P3 to win and loss (score at a single electrode) had excellent internal consistency (Chronbach's α) across 24 trials per condition and RewP to both win and loss at Fz and Pz had acceptable internal consistency across 12 trials per condition in a larger sample of 14–18-year-old adolescents oversampled for early depression (*custom pipeline*; Luking et al., 2021). In adults, the SPN, Cue P3, and RewP exhibited good to excellent internal consistency (Chronbach's α) across 25 trials per condition whereas the contingent negative variation (CNV) exhibited poor internal consistency (Oumeziane et al., 2019). No studies evaluated evidence of test–retest reliability of MID ERPs.

In case of the Doors task, regarding evidence of internal consistency, in children and young adolescents, the RewP and FN exhibited good internal consistency in a sample of 8–13-year-old girls (*custom pipeline*; Bress et al., 2015) and in a sample of 8–14-year-old girls (*custom pipeline*; Luking et al., 2017). The ΔRewP (as a difference *and* a residual score) exhibited poor internal consistency in 8-14-year-old girls (lower reliability was observed for the difference relative to the residual score) (Luking et al., 2017). Regarding test–retest reliability, in children and young adolescents (8–13-year-old girls), results show that controlling for age and depression, the RewP and FN have moderate to strong whereas the ΔRewP has relatively poor 2-week test–retest reliability (Bress et al., 2015). Similarly, the RewP and FN have moderate to strong rank-order temporal stability across late childhood through early adolescence to middle adolescence assessments, but the ΔRewP showed fair

to good test–retest reliability from early to middle adolescence (*custom pipeline*; Kujawa et al., 2018).

We are aware of no research on reliability of the RewP in middle-late adolescents. In adults, results show that both the RewP and FN exhibit good internal consistency across 20 trials for young (Bress et al., 2015; *custom pipeline*; Marco-Pallares et al., 2011) and across 50 trials for older adults (Marco-Pallares et al., 2011) as well as good 1-week test–retest reliability (Levinson et al., 2017). Lower internal consistency (across 20 trials) (Bress et al., 2015; Levinson et al., 2017) and 1-week test–retest reliability (Levinson et al., 2017) was observed for the ΔRewP. Good to excellent internal consistency was replicated in a broad age-range of 10–55-year-olds (*custom pipeline*; Ethridge & Weinberg, 2018).

1.1.2 | Validity

We are aware of no studies evaluating evidence of validity of the MID ERPs in children. In the single such study with adolescents, findings indicated that the Cue P3 to gain and loss and SPN to loss predicted less emotion dysregulation (construct validity) and the Target P3 to gain and loss predicted self-reported reward responsiveness (convergent validity) in 14–17-year-old youth (*using a custom preprocessing pipeline*; Zubovics et al., 2021). In adults, evidence supports construct validity of MID ERPs across several samples, indicating reward anticipation and reward response can be parsed (Novak & Foti, 2015) and also construct and convergent validity indexed by differential relations across components with depression and sensation-seeking (*custom pipeline*; Novak et al., 2016; *custom pipeline*; Novak & Foti, 2015).

In case of the Doors RewP, in children, findings of a number of studies are evidence of its construct and convergent validity (*custom pipeline*; Kujawa et al., 2014; Kujawa et al., 2019; Kujawa, Proudfit, Kessel, et al., 2015; Kujawa, Weinberg, et al., 2013), in relation to observed and self-reported positive affectivity, anxiety, and depression symptoms. In young adolescents, evidence supports cross-domain stability of the Doors RewP across monetary and social reward tasks, with youth exhibiting an equally large RewP to both reward types (*custom pipeline*; Ethridge et al., 2017). In middle-late adolescents, one study focused on validity of the Doors RewP. Data provided some, but not overwhelming evidence supporting construct validity, as both the RewP and FN predicted lower negative affectivity and less emotion dysregulation, and the FN predicted greater fight/flight/freeze system (FFFS) sensitivity (*custom pipeline*; Zubovics et al., 2021), as well as convergent validity, as the FN predicted self-reported reward responsiveness in 14–17-year-old youth (*custom pipeline*; Zubovics et al., 2021). In adults, evidence supports cross-domain stability of the RewP across monetary (Doors) and social reward tasks, with adults showing a larger RewP to monetary than to social reward (Ethridge et al., 2017), with evidence for cross-domain stability replicated in a broad age-range of 10–55-year-olds (Ethridge & Weinberg, 2018).

1.1.3 | Data quality

We are aware of no research on MID ERPs in children. In a sample of 15-year-old male adolescents and young adults, across Cue and Target P3, CNV, FRN, and the late positive potential (LPP), ERPs exhibited lower (albeit at the time considered acceptable) signal-to-noise ratio (SNR), with values ranging from 2.86 to 4.71 across 60 trials (Broyd et al., 2012). We are aware of no research on Doors RewP SNR in children or adolescents. In adults, in one experiment, SNR for the FRN was 5 for 30 trials and nearly 7 for 60 trials and in a second experiment, SNR was 3.5 for 30 trials and 4.5 for 40 trials (Marco-Pallares et al., 2011).

Taken together, despite increased attention devoted to assessing evidence of reliability (and validity) of ERPs, there is a paucity of research examining the psychometric properties of *ERP markers of reward processing—more specifically, of reward anticipation and reward response—in middle-late adolescents*, despite developmental differences being especially relevant during this phase.

1.2 | Current study

Our aim in the current research was to begin filling this gap in knowledge.

Previously, we examined and reported on questions of convergent and incremental validity of MID and Doors ERPs (Zubovics et al., 2021); in a sample of middle-late adolescents, we examined correspondence between ERP indices of reward anticipation and response and self-report measures of reinforcement sensitivity, whether ERP indices are related to individual differences in affectivity and affect regulation, and whether ERP indices predict these affective outcomes above and beyond self-report.

Here, our interest is in data quality and reliability. Accordingly, our first aim was to examine, in the same community sample of youth, the SNR, mean amplitude across trials, and evidence of internal consistency and of ERP indices of reward anticipation (Target P3, Cue P3, and feedback SPN) and reward response (RewP, FN, and ΔRewP).

For findings to inform conclusions regarding an approximation of the number of trials needed for stable and

internally consistent estimates, they need to be generalizable. As is true for an overwhelmingly large proportion of ERP research pipelines—including all MID and Doors reliability and validity studies reviewed under "Psychometric properties of MID and Doors ERPs," our laboratory also employs a custom data preprocessing pipeline. Artifact correction methods determine which trials are excluded and kept, they affect the time course and amplitude of the obtained ERPs which, in turn, affect all findings involving those ERPs, including assessments of psychometric properties. To circumvent difficulties with generalizability given differences in artifact correction methods and preprocessing pipelines, a number of standardized pipelines have been developed (Bigdely-Shamlo et al., 2015; Cowley et al., 2017; da Cruz et al., 2018; Debnath et al., 2020; Gabard-Durnam et al., 2018; Levin et al., 2018; Pedroni et al., 2019; Rodrigues et al., 2021). Our second aim was to compare findings (ERP time course and amplitude, SNR, stability, and internal consistency) obtained with our custom and a standardized (Debnath et al., 2020) preprocessing pipeline.

Finally, given the debate about whether or not the MID task is appropriate for measuring reward response (with most e-MID studies reporting the RewP, (Novak et al., 2016; Novak & Foti, 2015; Oumeziane et al., 2019), we assessed between-task and within-individual correspondence between MID and Doors RewPs.

2 | Method

2.1 | Procedure

Data were collected in the context of a larger longitudinal project, the Budapest Longitudinal Study of ADHD and Externalizing Disorders, aimed at identifying affective-motivational behavioral and biological protective and risk factors of behavior problems and functional impairments in adolescents. Data used in the current study were obtained during the first year of the larger project.

Participants between the ages of 14 and 18 years were recruited from public middle- and high schools in Budapest, Hungary. This research was approved by the National Institute of Pharmacy and Nutrition (OGYÉI/17089–8/2019) and has been performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. Participants' parents provided informed consent and participants provided assent. Following informed consent procedures, the EEG cap was applied and experimental tasks were administered. For the first half of the sample, the MID task followed by the Doors task and for the second half of the sample, the Doors task followed by the MID task (task order did not affect any of the ERP components, all false discovery rate-corrected *p*s > .35). (For waveforms, SNR and internal consistency findings given task order, see Supplementary Information Figures S1–S11). Following experimental procedures, participants were asked, in an unstructured interview, about their experiences regarding the paradigms, i.e., what, if anything, they enjoyed and what, if anything they did not enjoy and what else, if anything, they wished to add as pertinent feedback to the researchers.

2.2 | Participants

Participants were the same 43 adolescents (M_{age} = 15.67 years; *SD* = 1.01; range: 14–18 years; 32.6% boys) who participated in the first year of the larger project (a total of 51 youths participated and 43 has a sufficient number of trials and their data available for analyses), as in our prior study on convergent and incremental validity (Zubovics et al., 2021). All identified as Caucasian. Exclusionary criteria were (a) self-reported past or present diagnosis of any psychiatric or neurological disorder, such as pervasive developmental disorder, bipolar disorder, psychosis, substance dependence or epilepsy; (b) having visual impairment as defined by impaired vision <50 cm, unless corrected by glasses or contact lenses. One participant was excluded from all analyses because they were an outlier across many MID and Doors trials and one participant was excluded from all MID analyses because they reported they could not clearly see the amount of money shown indicated with a "+" or a "−" sign (i.e., they did not know whether they won or lost) during the paradigm (but did not report this until after completing the experimental paradigms). As such, 41 participants (M_{age} = 15.66 years; *SD* = 1.02; range: 14–18 years; 31.71% boys) were included in MID analyses and 42 participants (M_{age} = 15.69 years; *SD* = 1.02; range: 14–18 years; 30.95% boys) were included in Doors analyses.

Of note, it is reasonable to assume that the participant who was excluded from MID analyses could be retained for Doors analyses. In case of the MID task, there is written text to that is essential to be read for the paradigm to elicit its intended effects. In case of the Doors task, there is no such text. In case of the MID, the amount of money that can be won or lost is presented before each trial and so is the total amount of money won after each trial (in addition to the cue and target geometric shapes). Conversely, in case of the Doors task, there is no text that has to be read, as only the doors and the arrows are presented to participants. Nevertheless, to check whether Doors findings change with exclusion of this participant, we repeated all Doors analyses with this participant excluded.

Results did not meaningfully change (see Supplementary Information Tables S1 and S2).

2.3 | Experimental paradigms

As noted, a validated monetary incentive delay (MID) task was used to probe reward anticipation and initial responsiveness to reward attainment and the Doors task was used to probe initial responsiveness to reward attainment.

2.3.1 | Monetary incentive delay (MID) task

The MID task (Knutson et al., 2003; Knutson, Fong, Adams, et al., 2001; Knutson, Fong, Hommer, et al., 2001) consisted of 336 trials in total, presented in seven blocks of 12 trials/condition. During the task, participants responded to a sequence of geometric shapes indicating money (1000 Hungarian Forints [HUF]) can be gained (e.g., full circle, i.e., a win condition), or loss of money can be avoided (e.g., full square, i.e., a loss condition), or that it is a neutral trial (e.g., empty circle or square, i.e., two neutral conditions), with no monetary consequence. Following each cue (2000 ms duration), there was an anticipatory phase (duration between 2000 and 2500 ms). During the anticipatory phase, participants waited for and were briefly presented with a target stimulus that they had to quickly respond to with button press to gain or avoid losing money. Success or failure was indicated on the computer screen (2000 ms feedback duration), and so was the cumulative total money won. The duration of the intertrial interval was between 1000 and 2000 ms. The duration of the target stimulus was determined before the first block using a shorter training block. The target duration was set to a winning chance of 66%. Trials corresponding to different conditions were presented in a random order.

2.3.2 | Doors task

The Doors task (Dunning & Hajcak, 2007; Foti & Hajcak, 2009; Kujawa et al., 2014; Kujawa et al., 2018; Kujawa, Smith, et al., 2013) consisted of 180 trials in total, presented in three blocks of 30 trials/condition. Participants were told that on each trial, they could either gain 100 or lose 50 (HUF). At the beginning of each trial, a fixation mark (+) appeared for 900 ms. Then, participants were presented with an image of two doors for 3000 ms and asked to choose one door by pressing the number 7 or 8 on the keypad (for the left and the right door, respectively). Finally, after a short delay (1100 ms with a jitter of ±50 ms), feedback was presented for 1500 ms on the screen. Gain was indicated by a green "↑" and loss was indicated by a red "↓". The duration of the intertrial interval was 2000 ms with a jitter of ±250 ms. In a single block, 30 gain and 30 loss trials were presented in random order.

To maximize effectiveness of both paradigms, participants were told that the virtual money they accumulated during each task can be exchanged for fruits and snacks (candy, chips, etc.), with more virtual money exchangeable for more desirable fruit and snack options (as ranked by the participant prior to the tasks).

2.3.3 | EEG data acquisition and processing

EEG data were recorded and processed as described in our previous study (Zubovics et al., 2021). Briefly, continuous EEG was obtained at a sampling rate of 1000 Hz and digitized at 16-bit resolution with a 64-channel BrainAmp DC system equipped with actiCAP active electrodes (Brain Products GmbH, Gilching, Germany). The following electrodes were used to record the electrical activity of the brain: Fp1, Fp2, AF3, AF7, AF4, AF8, F1, F3, F5, F7, F2, F4, F6, F8, Fz, FC1, FC3, FC5, FT7, FT9, FC2, FC4, FC6, FT8, FT10, FCz, C3, C5, T7, C4, C6, T8, Cz, CP1, CP3, CP5, TP7, TP9, CP2, CP4, CP6, TP8, TP10, CPz, P1, P3, P5, P7, P2, P4, P6, P8, Pz, PO3, PO7, PO9, PO4, PO8, PO10, POz, O1, Oz, O2. The FCz electrode was used as online reference. Electrode impedances were kept below 15 kΩ and a 250 Hz hardware low-pass filter was applied on the recorded EEG. We also used two electrodes to record blinks and eye movements: electrooculogram (EOG) electrodes were placed (1) below the left eye and (2) lateral to the outer canthus of the right eye.

Offline processing of the EEG data was performed using the FieldTrip open source MATLAB toolbox (Oostenveld et al., 2011) and custom MATLAB analysis scripts (R2017a, The MathWorks, Inc, Natick, MA, USA). Hamming-windowed sinc finite impulse response (FIR) filters (passband deviation: 0.0022 [0.22%]; stopband attenuation: −53 dB) were used during the filtering steps of the EEG signal processing. These filters were implemented as one-pass zero-phase forward filters with delay compensation (built-in "firws" filter type in FieldTrip). Half-amplitude (−6 dB) cutoff frequencies are described here. We applied the following steps to preprocess the continuous EEG data (Zubovics et al., 2021): (1) Potential bad channels were detected using a custom MATLAB algorithm (sensitive to changes in the data due to large drifts and sudden voltage jumps), then, after a final visual inspection, removed or kept ($M \pm SD$: 0.53 ± 0.84 channels, range: 0–4). (2) Prior to independent component analysis (ICA)-based artifact rejection, we high-pass filtered the continuous EEG

at 1 Hz (order: 1650; transition width: 2 Hz; Winkler et al., 2015)). (3) Next, muscle artifact detection was performed on the filtered EEG using the automatic artifact rejection function of FieldTrip. Segments of EEG containing the detected muscle activity were marked for later removal. (4) ICA was conducted (logistic infomax ICA algorithm; Bell & Sejnowski, 1995) to remove blinks and eye movements. The topographical distribution and the time course of all ICA components were visually inspected to identify components representing EOG artifacts. In order to improve signal quality, occasionally other ICA components corresponding to transient or persistent noise artifacts were also marked for removal. (5) Selected ICA components were removed from the original, unfiltered EEG ($M \pm SD$: 3.45 ± 0.98 components, range: 1–8). (6) Then, we applied a 0.1 Hz high-pass filter (order: 16500; transition width: 0.2 Hz) on the ICA-cleaned EEG data. (7) After that, a weighted average of all neighboring channels of the same participant was used to interpolate bad channels previously removed. Channel weights were calculated based on the distances between the bad electrode and the surrounding electrodes. (8) Lastly, we re-referenced the preprocessed EEG data to the average of the electrodes located at the left and right mastoids (TP9 and TP10, respectively). Furthermore, the online reference electrode (FCz) was included in the group of active electrodes.

ERP averages were calculated from the preprocessed EEG (i.e., the final output of the preprocessing workflow) as follows (Zubovics et al., 2021). (1) The EEG was epoched from −200 ms (or, in case of the SPN, from −1200 ms) to 1000 ms around the stimuli (feedback, cue or target). (2) To ensure proper operation of our automatic artifact rejection algorithm, trials were low-pass filtered at 45 Hz (order: 294; transition width: 11.3 Hz). (3) Epochs containing high muscle activity (detected during step (3) of preprocessing) were removed. (4) An automatic artifact rejection method implemented in MATLAB was used to reject additional trials containing artifacts. Artifact removal was based on the following criteria: (i) a voltage step of more than 50 μV between data points, (ii) a voltage difference of 300 μV within a trial, and (iii) a voltage difference of <0.50 μV within 100 ms intervals (Bunford, Kujawa, Fitzgerald, et al., 2017; Bunford, Kujawa, Swain, et al., 2017; Kujawa et al., 2016; Kujawa, Proudfit, Kessel, et al., 2015). (5) We performed a final visual evaluation to detect and remove remaining epochs with artifacts (6) Next, trials were baseline corrected using the 200 ms time interval prior to the stimulus onset (for the SPN, the interval from −1200 ms to −1000 ms before the stimulus was used as baseline). (7) After that, for each participant and for each condition, we computed the ERP averages, then these averages were low-pass filtered at 30 Hz (order: 442; transition width: 7.5 Hz). (8) As a final step, for each component, grand average ERP waveforms were calculated from individual ERP averages. As such, based on chosen electrodes and time windows, one ERP value per condition was calculated for each participant.

Following artifact rejection, for each condition, participants had an average of 76.46 ± 12.30 trials (86.33% ± 11.80%, 6423/7440 trials in total) in case of the Doors task (range: 37–90, two participants had one less block, due to technical issues) and 73.69 ± 10.67 trials (90.24% ± 9.25%, 24,170/26784) in case of the MID task (range: 27–84, one participant had four less blocks, one participant had three less blocks and one participant had one less block due to technical issues).

Electrodes and time windows at which each ERP component was scored is as follows for the MID task: Cue P3: Pz, POz, P1, and P2, at 450–650 ms; Target P3: CPz, Pz, P1, and P2, at 200–375 ms; SPN: CPz, Pz, CP1, CP2, P1, and P2, at −200–0 ms; RewP, FN, and ΔRewP: CPz, Cz, FCz, CP1, CP2, FC1, and FC2, at 175–275. For the Doors task: RewP, FN, and ΔRewP: CPz, Cz, FCz, CP1, CP2, FC1, and FC2, at 175–275 ms.

Offline processing of EEG data was repeated using the Maryland analysis of developmental EEG (MADE) pipeline (Debnath et al., 2020). From the list of published standardized EEG preprocessing pipelines (Bigdely-Shamlo et al., 2015; Cowley et al., 2017; da Cruz et al., 2018; Debnath et al., 2020; Gabard-Durnam et al., 2018; Levin et al., 2018; Pedroni et al., 2019; Rodrigues et al., 2021) we chose the MADE pipeline because it was specially developed for the analysis of pediatric EEG data. For details, see Supplementary Information.

2.4 | Analytic plan

Data were analyzed using custom scripts in MATLAB (version 9.2 2017a).

For mean amplitudes across trials, we report findings as cumulative averages across all trials and, as 50% of participants had ≥74 trials, across the first 74 trials (Table 1). We also show results as grand average ERP waveforms across all trials (panel 1 of Figures 1–5). In reporting cumulative averages, we take into account the effect of trial count whereas in reporting grand averages, we follow convention. However, in case of grand averages, the corresponding values do not take into account the effect of trial count (i.e., the grand average value corresponding to a certain trial may reflect 40 participants' data or it may reflect, e.g., 20 participants' data, depending on the number of participants who had data for such number of trials). Findings are also presented as a function of increasing

TABLE 1 Descriptive statistics (in μVs) for all ERP components across the first 74 and across all trials

			First 74 trials[a]				All trials[b]			
					95% CI				95% CI	
	ERP	Condition	Mean	*SD*	Lower bound	Upper bound	Mean	*SD*	Lower bound	Upper bound
Doors	RewP	Win	7.88	4.55	6.45	9.32	7.84	4.40	6.45	9.23
	FN	Loss	7.16	4.58	5.71	8.60	7.08	4.53	5.65	8.51
	ΔRewP	Diff. score	0.73	1.85	0.14	1.31	0.76	1.89	0.16	1.35
MID	Cue P3	Win	9.27	4.14	7.96	10.57	9.35	4.19	8.03	10.67
	Cue P3	Loss	8.55	4.33	7.19	9.92	8.64	4.32	7.27	10.00
	Target P3	Win	10.14	4.33	8.77	11.50	10.13	4.33	8.76	11.49
	Target P3	Loss	9.91	4.49	8.49	11.33	9.89	4.48	8.48	11.31
	RewP	Win	9.22	3.41	8.15	10.30	9.16	3.41	8.09	10.24
	FN	Loss	8.91	3.35	7.85	9.97	8.85	3.36	7.79	9.91
	ΔRewP	Diff. score	0.31	1.42	−0.13	0.76	0.31	1.40	−0.13	0.75
	SPN	Win	−4.74	3.82	−5.94	−3.54	−4.72	3.79	−5.92	−3.53
	SPN	Loss	−5.32	4.09	−6.61	−4.02	−5.32	4.10	−6.61	−4.02

Note: Data are in μVs.

[a]Calculated by using the mean of the first 74 trials (or less) for every participant (does not equal the 74th data point on Figures 1–5).

[b]Calculated by using the mean of all trials for every participant (does not equal the final data point on Figures 1–5).

Abbreviations: CI, confidence interval; ERP, event-related potential; FN, feedback negativity, conceptualized here as a relative negativity following losses; MID, monetary incentive delay task; RewP, reward positivity, conceptualized here as a relative positivity following gains; SPN, stimulus preceding negativity; SD, standard deviation.

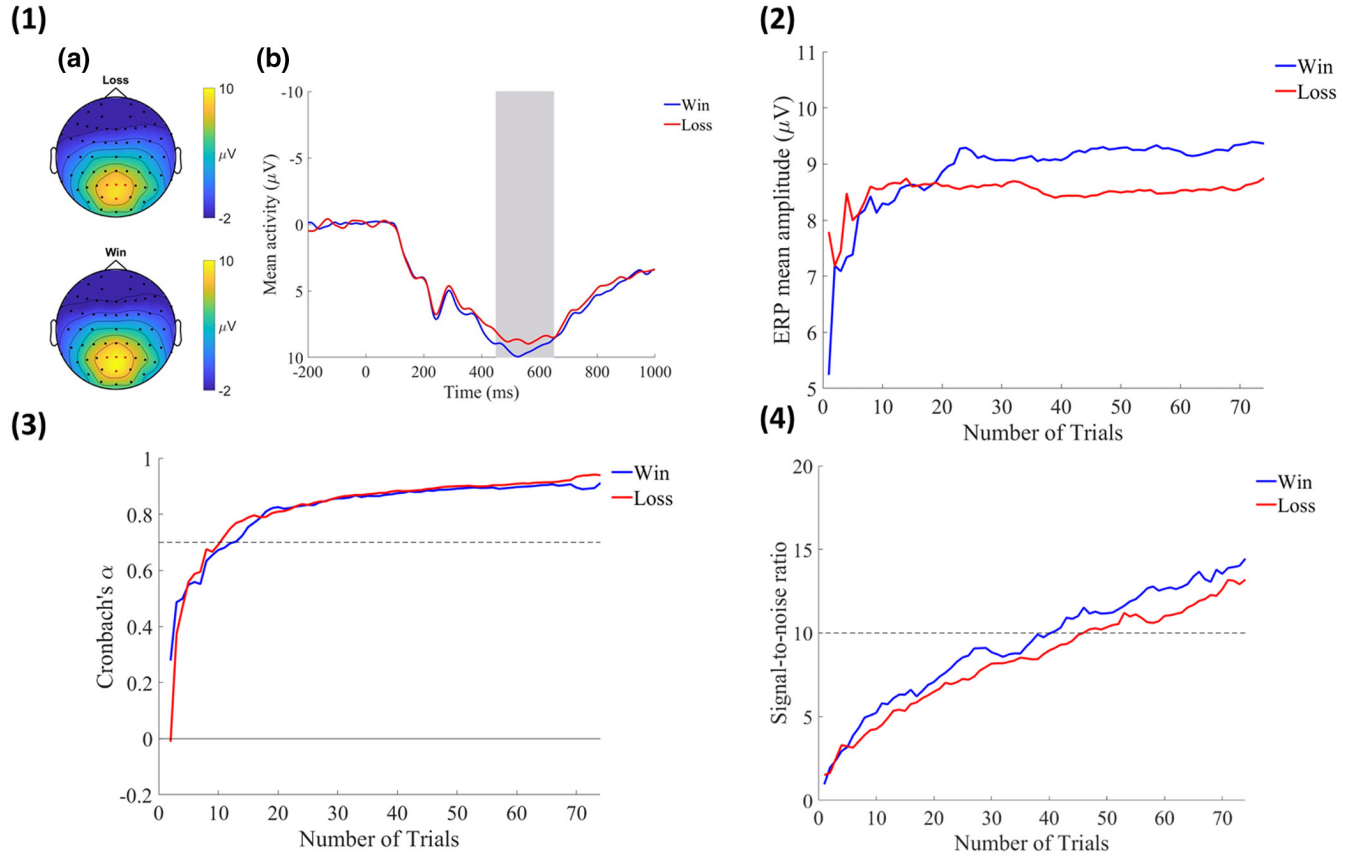

FIGURE 1 MID Cue P3. (1a) Scalp distributions depicting activation to Cue stimuli signaling win (Cue P3 win) and loss (Cue P3 loss) in the 450–650 ms time window, with electrodes selected for scoring the Cue P3 (Pz, POz, P1, and P2) in red. (1b) ERP grand average waveforms (*negative up*) of the win (blue) and loss (red) condition cues. Cue stimuli were presented at 0 ms and ERPs were scored in the 450–650 ms time window indicated by gray shading. (2) Mean activity for Cue stimulus (in μV) in the win (Cue P3 win; blue) and loss (Cue P3 loss; red) conditions as a function of the number of trials. (3) Internal consistency of the win (Cue P3 win; blue) and loss (Cue P3 loss; red) conditions, as measured using Cronbach's α. Threshold for an acceptable α (≥0.7) is indicated with a dashed line. (4) Signal-to-noise ratio (SNR) of the win (cue P3 win; blue) and loss (cue P3 lose; red) conditions as a function of the numbers of trials, with signal defined as the mean amplitude during of the component-relevant time window (450–650 ms) and noise defined as the mean of the absolute amplitude values during the baseline period (−200–0 ms). To calculate SNR, the signal was divided by the noise.

number of trials (Figures 1–5) to indicate the number of trials necessary for acceptable psychometric values for each ERP component.

Two indices of internal consistency were calculated for each ERP component of interest, that is, following Luking et al. (2017), the RewP reflecting a relative positivity following gains, the FN reflecting a relative negativity following losses, and the ΔRewP as the gain-loss difference score, as well as the Cue P3, Target P3, and SPN. (1) Split-half reliability conceptualized as the correlation between averages of even- and odd-numbered trials and corrected using the Spearman–Brown prophecy formula (Nunnally et al., 1967) and (2) Cronbach's α, conceptualized as roughly equivalent to the mean of all possible split-half correlations (Levinson et al., 2017). The advantage of the former is that it includes all available data (in either the average of even- or the average of odd-numbered trials) and its disadvantage is that it is specific to one way of splitting the data (i.e., even- vs. odd-numbered trials). Cronbach's α requires participants to have the same number of trials, resulting in data exclusion. All participants had ≥27 good gain and ≥28 loss trials for each ERP.

Internal consistency of the ΔRewP was estimated using an adjusted α formula (Lord, 1963). As the reliability of difference scores is affected by the reliabilities, variances, and intercorrelations of the two parent measures (here, the FN and RewP), standard reliability indices (i.e., split-half or Cronbach's α) are inappropriate measures for difference scores. In accounting for the reliabilities, variances, and intercorrelations of the parent measures, the adjusted α formula is a more accurate index of internal consistency (Furr & Bacharach, 2013).

SNR was calculated for each ERP component of interest using averages of increasing numbers of trials (trials 1 through 74 in steps of 1). To this end, first, for each

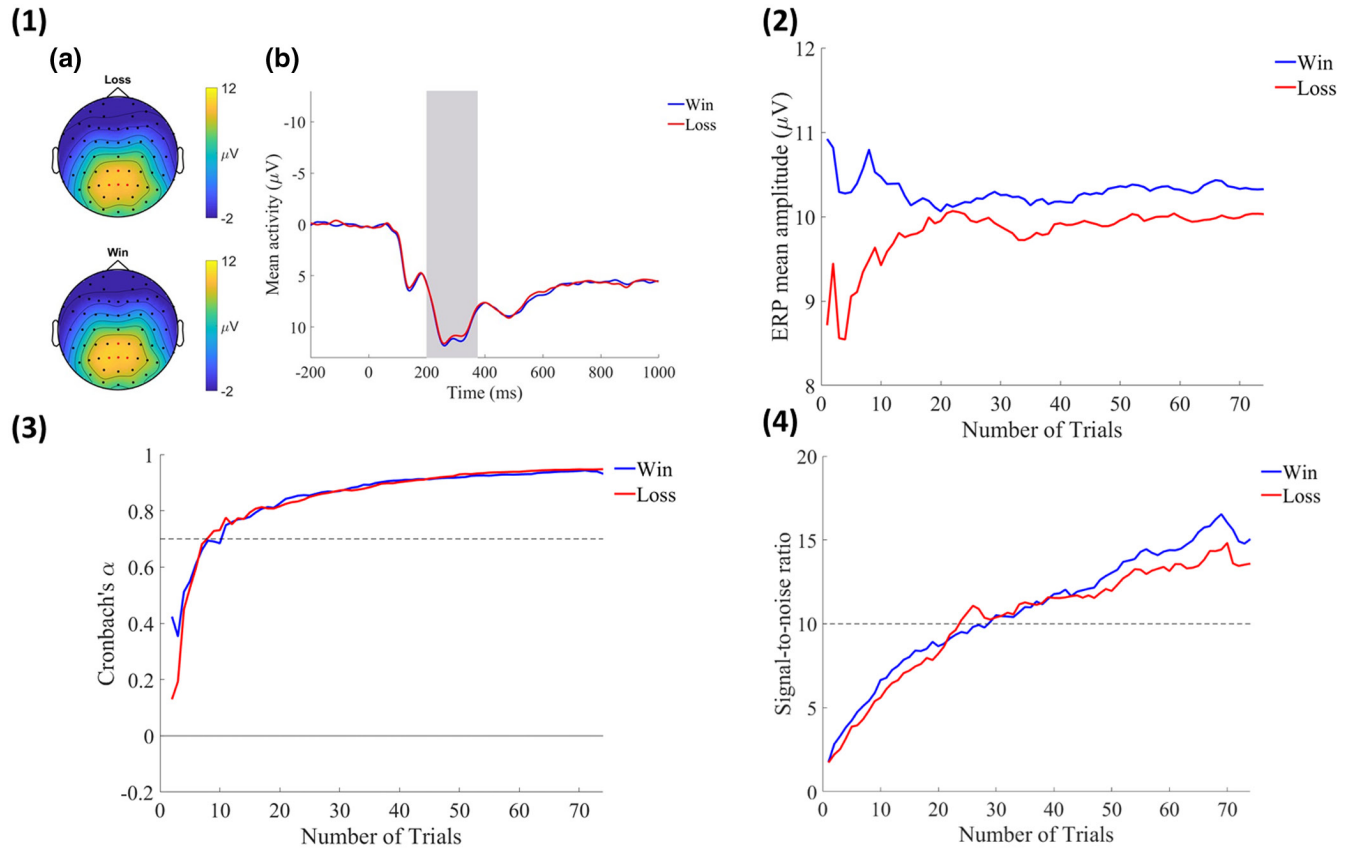

FIGURE 2 MID Target P3. (1a) Scalp distributions depicting activation to target stimuli signaling win (Target P3 win) and loss (Target P3 loss) in the 200–375 ms time window, with electrodes selected for scoring the Target P3 (CPz, Pz, P1, and P2) in red. (1b) ERP grand average waveforms (*negative up*) of the win (blue) and loss (red) condition cues. Target stimuli were presented at 0 ms and ERPs were scored in the 200–375 ms time window indicated by gray shading. (2) Mean activity for target stimulus (in μV) in the win (Target P3 win; blue) and loss (Target P3 loss; red) conditions as a function of the number of trials. (3) Internal consistency of the win (Target P3 win; blue) and loss (Target P3 loss; red) conditions, as measured using Cronbach's α. Threshold for an acceptable α (≥0.7) is indicated with a dashed line. (4) Signal-to-noise ratio (SNR) of the win (Target P3 win; blue) and loss (Target P3 lose; red) conditions as a function of the numbers of trials, with signal defined as the mean amplitude during of the component-relevant time window (200–375 ms) and noise defined as the mean of the absolute amplitude values during the baseline period (−200–0 ms). To calculate SNR, the signal was divided by the noise.

participant, SNRs were calculated by dividing the mean amplitude of the component-relevant time window with the mean of the absolute amplitude across the baseline (−200 to 0 ms for all components except for the SPN, where baseline was −1200 to −1000 ms). Second, SNRs were averaged across participants.

For SNR and reliability across trials, we present results across the first 74 trials, as a function of increasing number of trials (Figures 1–5).

All analyses were conducted with the custom pipeline-processed data and repeated with the MADE pipeline-processed data. For exploratory analyses, to assess between-task correspondence between MID and Doors RewP, MID and Doors RewP to win, RewP to lose, and ΔRewP values were compared in partial correlations (controlling for effects of task order). To assess within-person correspondence between MID and Doors RewP, MID and Doors RewP to win, RewP to lose, and ΔRewP values were compared in repeated measures ANCOVAs (controlling for effects of task order). α was considered acceptable if >0.7 and SNR was considered acceptable if ≥10[2] (Luck, 2014).

[2]SNR is calculated differently across studies (and this is acceptable; Luck, 2014), for example, with some authors using a single peak (Thigpen et al., 2017), others using peak-to-peak (Luck, 2014), and yet others using fast fourier transform (Boudewyn et al., 2018). Accordingly, we suggest the SNR ≥10 as the threshold for acceptability should be applied and interpreted with caution. (Others suggest a threshold may not even be valuable, as importance of a large SNR value depends on effect size and sample size. Thus, it is more reasonable for researchers to focus, rather than on a specific threshold, on decreasing noise and increasing signal as much as possible; Luck, 2014).

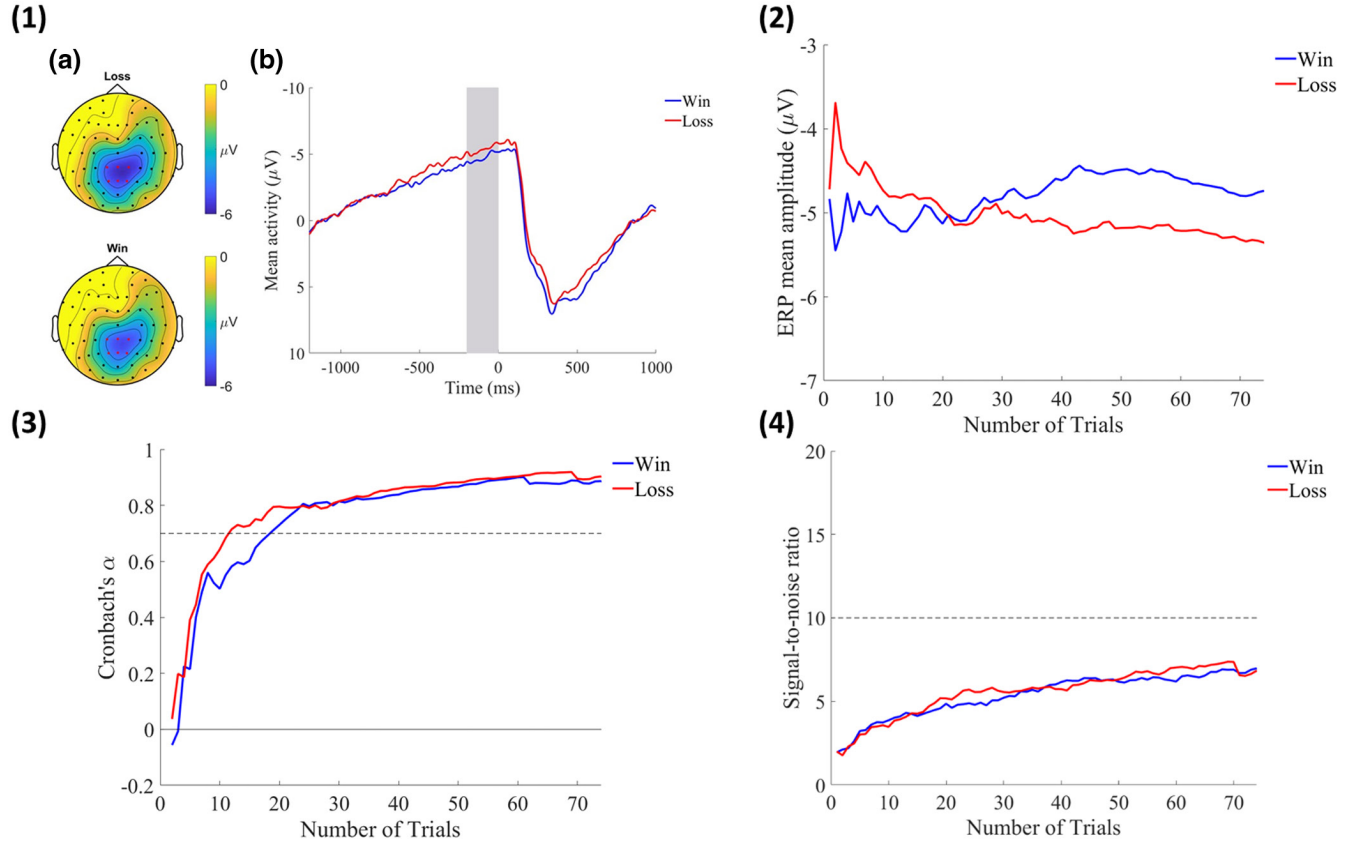

FIGURE 3 MID SPN. (1a) Scalp distributions depicting activation before feedback in win (SPN win) and loss (SPN loss) in the −200–0 ms time window, with electrodes selected for scoring the SPN (CPz, Pz, CP1, CP2, P1, and P2) in red. (1b) ERP grand average waveforms (*negative up*) of the win (blue) and loss (red) condition cues. Feedback stimuli were presented at 0 ms and ERPs were scored in the −200–0 ms time window indicated by gray shading. (2) Mean activity before feedback stimulus (in μV) in the win (SPN win; blue) and loss (SPN loss; red) conditions as a function of the number of trials. (3) Internal consistency of the win (SPN win; blue) and loss (SPN loss; red) conditions, as measured using Cronbach's α. Threshold for an acceptable α (≥0.7) is indicated with a dashed line. (4) Signal-to-noise ratio (SNR) of the win (SPN win; blue) and loss (SPN lose; red) conditions as a function of the numbers of trials, with signal defined as the mean amplitude during of the component-relevant time window (−200–0 ms) and noise defined as the mean of the absolute amplitude values during the baseline period (−1200–1000 ms). To calculate SNR, the signal was divided by the noise.

3 | RESULTS

For grand averages and scalp distributions across ERP components of interest, see panel 1 of Figures 1–5.

3.1 | Stabilization

As indicated by visual inspection, the mean amplitude of the Cue P3 to gain increased somewhat sharply until the 25th trial, after which it stabilized (Figure 1). The mean amplitude of the Cue P3 to loss increased gradually until the 8th trial, after which it stabilized (Figure 1). The mean amplitude of Target P3 to gain decreased, and to loss increased, until the 22nd trial, after which both stabilized (Figure 2). In case of SPN to gain, visual inspection showed its amplitude gradually, slightly increased until the 40th trial, after which it stabilized. In case of SPN to loss, its amplitude decreased somewhat more sharply until the 20th trial, after which it stabilized (Figure 3). The mean amplitude of both MID RewP and MID FN sharply decreased within the first 10 and five trials, respectively, then gradually and slightly increased, and then very slightly decreased slightly to stabilize after the 20th trial, whereas the mean amplitude of the MID ΔRewP remained roughly consistent beyond the 14th trial, throughout the task (Figure 4).

The mean amplitudes of Doors RewP and Doors FN gradually stabilized over the course of the task, leveling off but continuing to gradually, slightly decrease after 20 trials, whereas the mean amplitude of the Doors ΔRewP remained roughly consistent beyond the 20th trial, throughout the task (Figure 5).

See Table 1 for descriptive statistics for all ERP components across the first 74 and across all trials and grand averages.

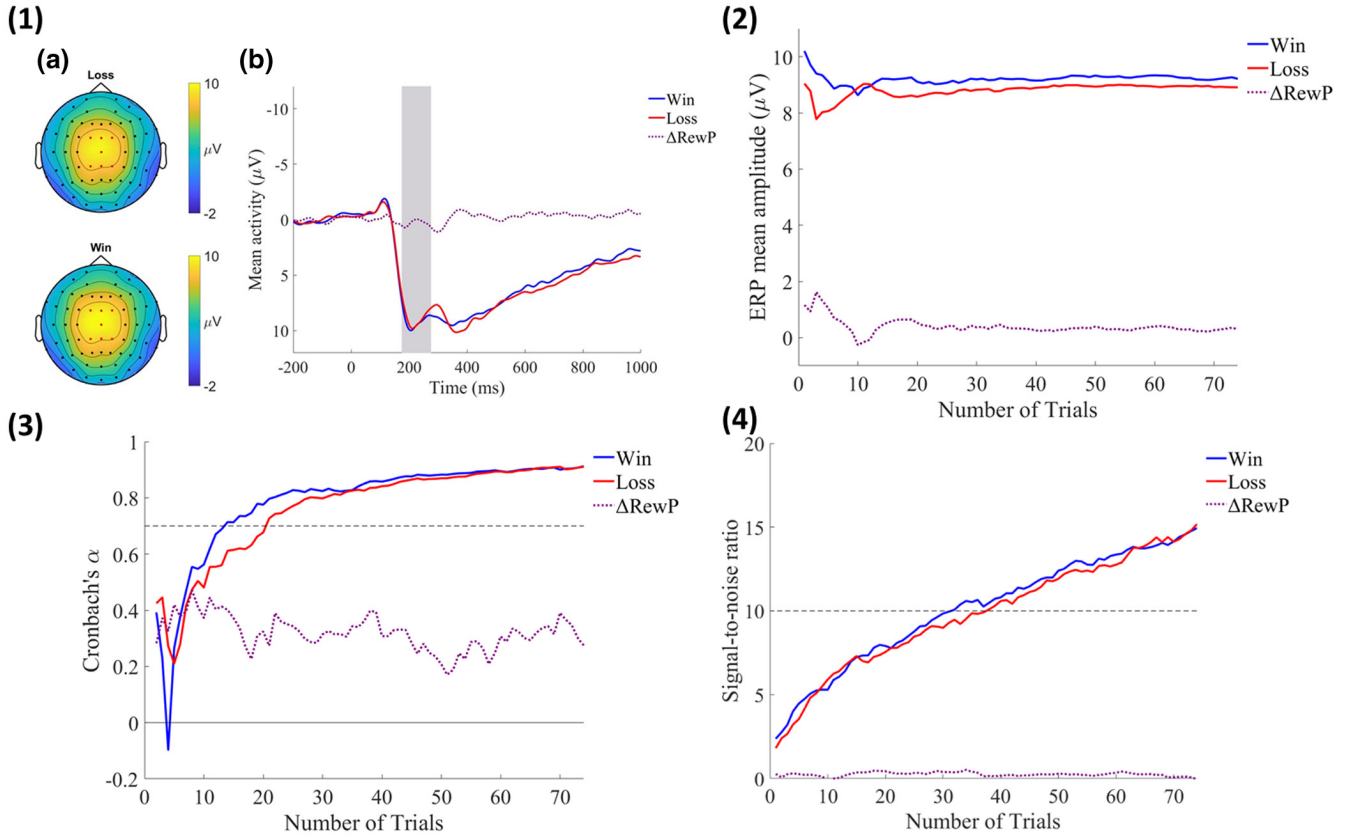

FIGURE 4 MID RewP. (1a) Scalp distributions depicting activation to feedback in win (MID RewP), loss (MID FN), and the difference score for win-loss (MID ΔRewP) in the 175–275 ms time window, with electrodes selected for scoring (CPz, Cz, FCz, CP1, CP2, FC1, and FC2) in red. (1b) ERP grand average waveforms (*negative up*) of the win (MID RewP; blue) and loss (MID FN; red) condition cues as well as the difference score (MID ΔRewP; purple). Feedback stimuli were presented at 0 ms and ERPs were scored in the 175–275 ms time window indicated by gray shading. (2) Mean activity to feedback stimulus (in μV) in the win (MID RewP; blue) and loss (MID FN; red) conditions as well as the difference score (MID ΔRewP; purple) as a function of the number of trials. (3) Internal consistency of the win (MID RewP; blue) and loss (FN; red) conditions, as measured using Cronbach's α and the internal consistency of the difference score (MID ΔRewP; purple) as measured using adjusted alpha (Lord, 1963). Threshold for an acceptable α (≥0.7) is indicated with a dashed line. (4) Signal-to-noise ratio (SNR) of the win (MID RewP; blue) and loss (MID FN; red) conditions as well as the difference score (MID ΔRewP; purple) as a function of the numbers of trials, with signal defined as the mean amplitude during of the component-relevant time window (175–275 ms) and noise defined as the mean of the absolute amplitude values during the baseline period (−200–0 ms). To calculate SNR, the signal was divided by the noise.

3.2 | Internal consistency

3.2.1 | Split-half reliability and Cronbach's alpha values across 74 trials[3]

The Cue P3 to both gain and loss achieved good to excellent internal consistency as assessed using both split-half reliability (*r*s 0.87 and 0.92, respectively) and Cronbach's α (0.91 and 0.94, respectively). The Target P3 to both gain and loss achieved excellent internal consistency as assessed using both split-half reliability (*r*s 0.93 and 0.95, respectively) and Cronbach's α (0.93 and 0.95, respectively). The SPN to both gain and loss achieved good to excellent internal consistency as assessed using both split-half reliability (*r*s 0.94 and 0.93, respectively) and Cronbach's α (0.89 and 0.90, respectively). The MID RewP and MID FN achieved good to excellent internal consistency as assessed using both split-half reliability (*r*s 0.94 and 0.94, respectively) and Cronbach's α (0.91 and 0.91, respectively). Across 74 trials, internal consistency of the MID ΔRewP was unacceptable (adjusted $\alpha = 0.27$).

The Doors RewP and Doors FN achieved excellent internal consistency as assessed using both split-half reliability (*r*s 0.93 and 0.95, respectively) and Cronbach's α (0.94 and 0.94, respectively). Across 74 trials, internal consistency of the ΔRewP was unacceptable (adjusted $\alpha = 0.03$), though across 72 trials it was, albeit unacceptable, somewhat higher (adjusted $\alpha = 0.37$).

[3]Values reported are those observed at the 74th trial.

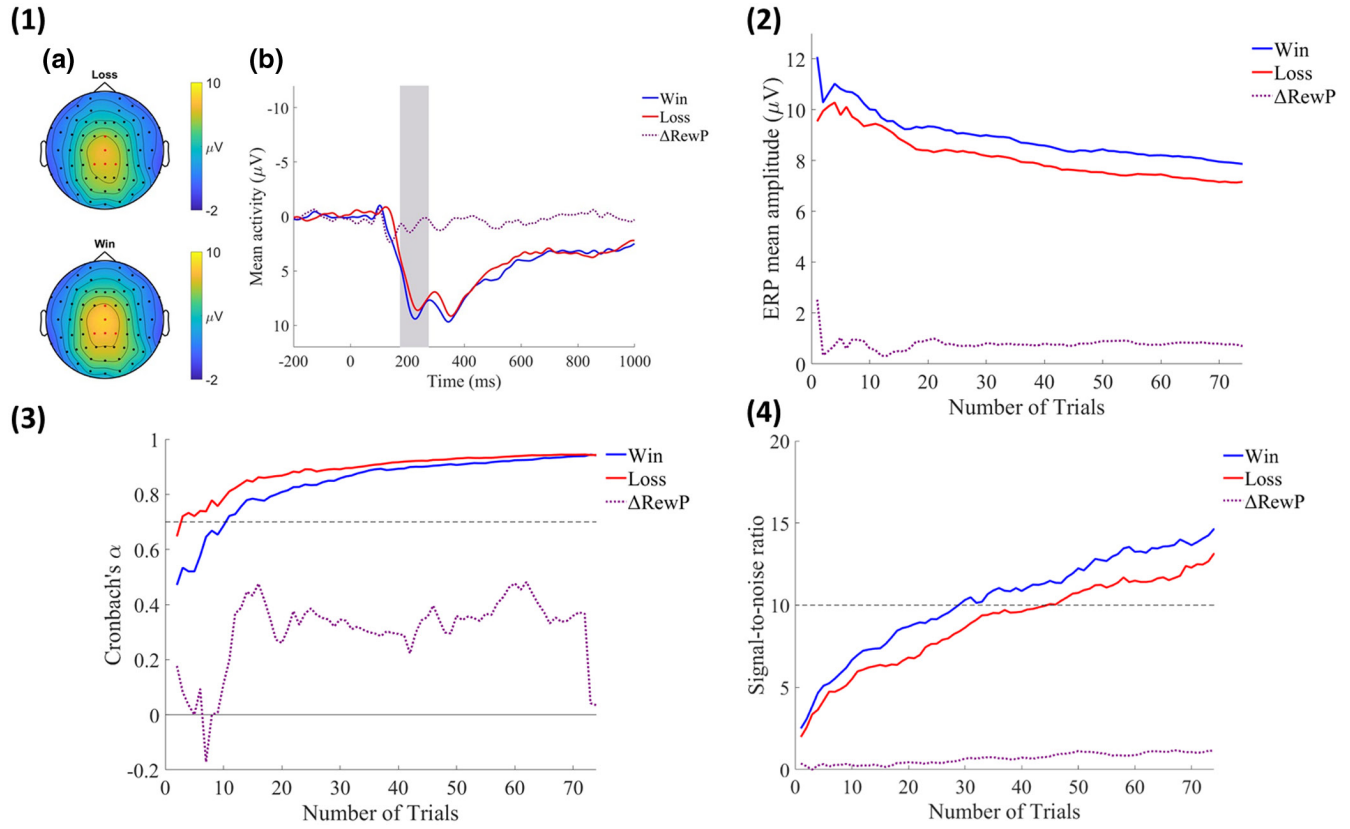

FIGURE 5 Doors RewP. (1a) scalp distributions depicting activation to feedback in win (doors RewP), loss (doors FN), and the difference score for win-loss (doors ΔRewP) in the 175–275 ms time window, with electrodes selected for scoring (CPz, Cz, FCz, CP1, CP2, FC1, and FC2) in red. (1b) ERP grand average waveforms (*negative up*) of the win (doors RewP; blue) and loss (doors FN; red) condition cues as well as the difference score (doors ΔRewP; purple). Feedback stimuli were presented at 0 ms and ERPs were scored in the 175–275 ms time window indicated by gray shading. (2) Mean activity to feedback stimulus (in μV) in the win (doors RewP; blue) and loss (doors FN; red) conditions as well as the difference score (doors ΔRewP; purple) as a function of the number of trials. (3) Internal consistency of the win (doors RewP; blue) and loss (FN; red) conditions, as measured using Cronbach's α and the internal consistency of the difference score (doors ΔRewP; purple) as measured using adjusted alpha (Lord, 1963). Threshold for an acceptable α (≥0.7) is indicated with a dashed line. (4) Signal-to-noise ratio (SNR) of the win (doors RewP; blue) and loss (doors FN; red) conditions as well as the difference score (doors ΔRewP; purple) as a function of the numbers of trials, with signal defined as the mean amplitude during of the component-relevant time window (175–275 ms) and noise defined as the mean of the absolute amplitude values during the baseline period (−200–0 ms). To calculate SNR, the signal was divided by the noise.

3.2.2 | Cronbach's alpha values as a function of the number of trials

Acceptable internal consistency ($\alpha \geq 0.7$) was reached and maintained for the: Cue P3 to gain by the 12th, Cue P3 to loss by the 11th (Figure 1), Target P3 to gain by the 11th, Target P3 to loss by the 8th (Figure 2), and SPN to gain by the 19th, and SPN to loss by the 12th trial (Figure 3). Acceptable internal consistency ($\alpha > 0.7$) was reached and maintained for the MID RewP by the 14th and the MID FN by the 21st trial. In case of the MID ΔRewP, acceptable internal consistency was never achieved, not even at trial 8, which had the largest adjusted alpha ($\alpha = 0.47$) (Figure 4).

Acceptable internal consistency ($\alpha \geq 0.7$) was reached and maintained for the Doors RewP by the 11th and for the Doors FN by the third trial. In case of the ΔRewP, acceptable internal consistency was never achieved, not even at trial 16 (and 60 and 62), which had the largest adjusted alpha ($\alpha = 0.48$) (Figure 5).

Before internal consistency analyses were conducted, on average, 1.57 (range: 0–10) trials were removed because of artifacts for the MID and 0.76 (range: 0–10) for the Doors paradigm.

3.3 | Signal-to-noise ratio

Visual inspection indicated SNR linearly increased as a function of logarithmically increasing number of trials. Acceptable SNR values (SNR ≥10 (Luck, 2014)) were reached and maintained, for the Cue P3 to gain and loss

within 41 and 38 trials (and were 14.44 and 13.49 across 74 trials, i.e., at the 74th trial), respectively (Figure 1) and the Target P3 to gain and loss within 29 and 24 trials (and were 15.05 and 13.59 across 74 trials), respectively (Figure 2). Acceptable SNR values were never reached in case of SPN to gain or loss, where the SNR highest values were 6.97 and 7.37 (and were 6.98 and 6.84 across 74 trials), respectively (Figure 3). Acceptable SNR values were reached and maintained, for the MID RewP and MID FN within 38 and 32 trials (and were 15.16 and 14.93 across 74 trials), respectively (Figure 4).

Acceptable SNR values were reached and maintained, in case of the Doors RewP, within 29 and in case of the Doors FN, within 45 trials (and were 14.65 and 13.15 across 74 trials, respectively) (Figure 5).

Before SNR analyses were conducted, on average, 3.50 (range: 0–20) trials were removed because of artifacts for the MID and 4.74 (range: 0–19) for the Doors paradigm.

3.4 | Exploratory analyses

Partial correlations (controlling for effects of task order) indicated both RewP to win ($r = 0.604$, $p < .001$) and RewP to lose ($r = 0.726$, $p < .001$) are correlated across tasks, whereas ΔRewP values are not ($r = 0.141$, $p = .386$). Repeated measures ANCOVAs (controlling for effects of task order) indicated MID and Doors RewP to win values were different within-individuals, $F(1, 39) = 12.088$, $p = .001$ ($\eta_p^2 = 0.237$), with covariate-adjusted means showing the MID elicits a larger RewP to win ($M_{\text{MIDRewPwin}} = 9.165$, SE = 0.666; $M_{\text{DoorsRewPwin}} = 7.909$, SE = 0.537). MID and Doors RewP to lose values were also different within-individuals, $F(1, 39) = 9.680$, $p = .003$ ($\eta_p^2 = 0.199$), with covariate-adjusted means showing the MID elicits a larger RewP to lose ($M_{\text{MIDRewPwin}} = 8.850$, SE = 0.530; $M_{\text{DoorsRewPwin}} = 7.110$, SE = 0.708). MID and Doors ΔRewP values were not different within individuals, $F(1, 39) = 1.372$, $p = .249$ ($\eta_p^2 = 0.034$).

4 | DISCUSSION

Our aim in this research was to assess evidence of the reliability, specifically, the internal consistency, as well as mean amplitude across trials and SNR of ERP components of reward anticipation, namely the Cue P3, the Target P3, and the SPN, and of initial responsiveness to reward attainment, namely the Rewp, FN, and ΔRewP. Henceforth, we discuss our findings and make recommendations for the necessary and sufficient number of trials for internally consistent and stable estimates for these ERP components in middle-late adolescents. Of note, our recommendations are just that—suggestions based on results obtained in our specific design with our specific sample, acknowledging that as ERP psychometrics vary widely across studies, definitive conclusions about necessary number of trials or overall reliability of these measures cannot be drawn based on any one study. Nevertheless, as we employed a custom and a standardized preprocessing pipeline and obtained comparable values across the two, our results—and thus conclusions and recommendations for necessary and sufficient number of trials—are independent of our specific artifact correction method/ subjective parameters and can arguably be generalized under the specifications of standardization.

With regard to reward anticipation, this was the first attempt at examining evidence of the reliability of relevant ERP components as probed by the MID task. Findings indicate most components were stable after 25 trials and all were stable by the 40th trial (greatest number of trials needed for stable estimates was for SPN to gain). All components achieved good to excellent internal consistency across 74 trials as assessed using split-half reliability and Cronbach's α. Most achieved acceptable internal consistency after 14 trials, and all achieved acceptable internal consistency by the 19th trial (greatest number of trials needed for acceptable internal consistency was for SPN to gain). Most components achieved acceptable SNR after 38 trials and all achieved such threshold by 41 trials (greatest number of trials needed was for Cue P3 to gain), with the exception of the SPN, which did not achieve acceptable SNR within 74 trials. Taken together, our results show that in the current design and middle-late adolescent sample, in case of the ERP components of reward anticipation examined here, 41 trials per condition were necessary and sufficient to achieve stable and internally consistent estimates with acceptable SNR. Importantly, in case of SNR, on average, 3.50 pre-acceptable-level artifacted trials were removed, indicating *the actual number of necessary trials was 45*.

Of note, the SPN (especially to gain) stabilized later than the other herein assessed ERP components and did not achieve acceptable SNR. As observable on the figure depicting the grand average SPN waveforms (Figure 3), it is apparent that the slowly growing negativity already begins during the baseline period, making it noisy. As SNR is a function, in part, of the baseline amplitude, noisiness of the baseline will necessarily negatively affect SNR. In turn, a lower SNR will necessarily negatively affect the latency of stabilization. Taken together, these considerations indicate the baseline period for the SPN may need to be moved earlier so as to have a less noisy baseline for this component. In our case, this would not have been possible as doing so would have necessitated that we shorten an already almost too brief feedback time window.

With regard to initial responsiveness to reward attainment, this was the first attempt at examining evidence of the reliability of relevant ERP components as probed by the MID task and the first attempt at examining such evidence for ERP components probed by the Doors task in middle-late adolescents. Our data indicate that both RewP and the FN, in both the MID and the Doors task, were stable by the 20th trial. All components achieved good to excellent internal consistency across 74 trials as assessed using split-half reliability and Cronbach's α. Combined, MID RewP and FN achieved acceptable internal consistency after the 21st and Doors FN and RewP achieved acceptable internal consistency after the 11th trial. Combined, MID RewP and FN achieved acceptable SNR after the 38th and Doors RewP and FN achieved acceptable SNR after the 45th trial. These results suggest that in the current design and middle-late adolescent sample, in case of the ERP components of initial response to reward examined here, 38 (in case of the MID task)–45 (in case of the Doors task) trials per condition were necessary and sufficient to achieve stable and internally consistent estimates with acceptable SNR. Of note, in case of SNR, on average, 3.50 (in case of the MID task) and 4.74 (in case of the Doors task) pre-acceptable-level artifacted trials were removed, indicating *the actual number of necessary trials was 42 and 50*, respectively. This number is somewhat higher than that recommended by others and typically used in the literature with children and adults. Specifically, others have found that both the RewP and the FN achieved good to excellent internal reliability within 20 (Bress et al., 2015; Levinson et al., 2017; Marco-Pallares et al., 2011) to 30 trials (Luking et al., 2017) in the Doors task (with the exception of one study with older adults suggesting need for 50 trials; Marco-Pallares et al., 2011). Further, across studies where the Doors task is used to probe initial response to reward, typically, a total of 60 trials (30 per condition) (Bunford et al., 2021; Kujawa et al., 2014; Kujawa, Proudfit, Kessel, et al., 2015; Kujawa, Smith, et al., 2013) are used to generate ERPs. Importantly, when making recommendations as to the number of trials needed, to the best of our knowledge, others have not taken into account the number of trials that had artifacts and thus had to be removed, indicating the current recommendations for children and adults may be an underestimate.

Across components examined, it was only in case of the initial response to reward ERPs probed by the Doors (but not the MID) task, whose amplitude decreased over time (Figure 5). Across participants, many indicated in response to questions of our unstructured interview, that they found the Doors task boring or uninteresting. This might explain the observed amplitude decrease in the components probed by this task and highlights that there is a balance to be found with regard to having enough trials for stable and reliable estimates while also not having an overly large number of trials that no longer capture participants' attention and, as such, become invalid indices of the processes intended to be probed by the paradigm.

The herein evaluated psychometric properties of the ΔRewP, calculated as a difference score and probed in both the MID and the Doors task, were notably worse than for either of its parent measures (i.e., the RewP and the FN). The mean amplitude of the ΔRewP, albeit lower than observed in other studies (e.g., Kujawa et al., 2019), remained consistent throughout both tasks. Across 74 trials, internal consistency of the MID ΔRewP and the Doors ΔRewP was unacceptable and, although in both the MID and the Doors tasks, there were individual trials at which the adjusted alpha value was bordering poor, acceptable internal consistency was never achieved. These findings are generally consistent with prior results also showing internal consistency of the ΔRewP was unacceptable (Bress et al., 2015; Levinson et al., 2017). The obtained adjusted alpha values were comparable across the current and earlier studies (e.g., values of 0.37 and 0.27 across the MID and the Doors tasks, respectively, in the current research and 0.284 and 0.375 across assessments in Levinson et al., 2017). We echo the explanatory hypotheses offered by others, namely, that high intercorrelations between parent measures, unequal variances of parent measures, or poor reliability of the parent measures may each adversely affect reliability of the difference score. In the current study, the first two of these factors are true, suggesting high intercorrelations between – and unequal variance of—parent measures likely contributed to ΔRewP exhibiting unacceptable internal consistency.

Related, difference scores have been criticized (De Los Reyes, 2017; Laird & De Los Reyes, 2013) including for not meaningfully contributing incremental or unique information about, that is, being redundant with, the measures used to create them (Edwards, 1994). This issue poses a major challenge in affective/cognitive neuroscience and physiological research, where there is need to account for changes across conditions in case of certain processes of interest. Indeed, poor reliability of difference scores is observable in case of measures other than ERPs; a meta-analysis of 90 functional MRI (fMRI) experiments revealed poor reliability of task-fMRI measures, for example, an average test–retest reliability coefficient across over 1000 participants of 0.397, which is well below the cutoff for clinical applicability (ICC ≥0.8) and for individual-level interpretation (ICC ≥0.9) (Elliott et al., 2020). Elliott and colleagues argued that such poor reliability of fMRI measures is not due to poor reliability of MRI measures, or even of the BOLD signal itself; in the same meta-analysis, structural MRI measures evinced high test–retest reliability (Elliott et al., 2020) and in other studies, so did intrinsic

functional connectivity, that is, resting state MRI measures (Elliott et al., 2019). As such, rather than the tool itself being problematic, it is the adoption of approaches that are ideal for experimental cognitive neuroscience, that is, approaches that rely on differences in behavioral response to task vs. control conditions, that is, difference scores—that appear to be problematic for reliably measuring differences in neural activation across individuals. An alternative to difference scores is conducting analyses with residual scores though as noted, the ΔRewP calculated as a difference and as a residual score has been shown to exhibit poor internal consistency (Luking et al., 2017). Regardless of what the resolution to this issue will be, when considering what internal consistency represents, findings across studies suggest that difference scores are more constrained with regard to the amount of true score that is available in them. A relevant issue then becomes the amount of reliable variance that is available to relate to (e.g., clinical) outcomes. In the case of the ΔRewP, despite having potentially less true score variance than its parent measures, it has been relatively consistently shown to relate as well or better than the RewP or the FN to individual differences in depression (Levinson et al., 2017). These considerations underscore the importance, as also recommended by others (Levinson et al., 2017), of reporting for each dataset and sample, the psychometric properties (i.e., internal consistency) of ERPs, especially in studies where relations between ERPs and other measures of individual differences are examined.

Obtained peak amplitudes and time courses of ERPs were almost identical with our custom EEG data preprocessing pipeline and the standardized MADE pipeline, though amplitudes obtained with MADE were slightly smaller, likely because MADE has a less conservative artifact rejection procedure (MID trials retained with our custom pipeline: 73.69 ± 10.67 vs. with MADE: 79.06 ± 13.23 and Doors trials retained with our custom pipeline: 76.46 ± 12.30 trials vs. with MADE 82.64 ± 11.75 trials). Differences between the pipelines with regard to stability were variable; in case of some ERPs, stability was reached earlier (9 trials earlier for SPN to gain with MADE), in case of others, it was reached later (30 trials later for both Target P3s), and in case of yet others, it was reached after approximately the same number of trials (Cue P3, MID and Doors RewP, FN, and ΔRewP). In case of reward anticipation components, the amplitude at which components stabilized was somewhat lower with the MADE relative to the custom pipeline, though where differences were observed, those were negligible, usually 0.5 μV and never larger than 1 μV. In case of reward receipt components, the amplitude at which components stabilized was comparable or the same across pipelines. Of note, the mean amplitude of pre-stabilization trials was more variable with the MADE pipeline. Acceptable internal consistency was reached after a comparable number of trials and with nearly identical alpha values (negligibly lower with MADE), with the exception of both MID and Doors ΔRewP, which had a considerably better (0.27 vs 0.53 and 0.03 vs 0.40, respectively) – albeit still unacceptable—adjusted alpha value with MADE. Acceptable SNR was reached, however, after a greater number of trials with MADE (range = 6–17) (exceptions are MID and Doors RewP and FN; for these components, acceptable SNR was reached at the same time or, in case of Doors FN, 6 trials earlier with MADE), but 74-trial SNR values were comparable across pipelines. Taken together, the peak amplitudes and time courses of ERPs as well as the examined psychometric properties of those ERPs were comparable across our custom and the standardized preprocessing pipeline. Where differences were observed, those were minor or negligible. Arguably, greater leniency in artifact rejection can explain all observed differences: as less trials are removed, there is greater variability in the data and this necessitates a greater number of trials to achieve comparable psychometric properties as with a conservative artifact rejection.

As discussed, although neuroimaging measures have been increasingly used in individual differences and clinical research, research evaluating the psychometric properties of such measures has lagged behind and thus the utility of functional neuroimaging in translational research and clinical practice was even recently up for debate (Ball et al., 2014). An important area where neuroimaging has the potential to impact clinical care is in informing individualization of intervention, i.e., in informing single patient-level prediction of prognosis to guide clinical decision-making (Ball et al., 2014). For this, functional neuroimaging has to exhibit evidence of reliability as well as clear incremental validity (over easier and less costly measures), replicability (including in independent samples), and good predictive performance (Ball et al., 2014). Here, we present evidence of acceptable reliability of ERP measures of reward anticipation and initial responsiveness to reward attainment, with evidence linking the former to transdiagnostic characteristics such as affectivity and affect regulation (Zubovics et al., 2021) and the latter to various forms of psychopathology, including anxiety, depression, and ADHD (Bress et al., 2012; Bress et al., 2013; Bunford et al., 2021; Kessel et al., 2015). The current findings indicate, for the first time, that the Cue P3, Target P3, and SPN have good to excellent psychometric properties in middle to late adolescence in terms of both internal consistency and, with the exception of the SPN, SNR. The current results also extend earlier evidence underscoring reliability of the RewP and FN in adults and children/ younger adolescents, by showing that these

components also have good to excellent psychometric properties (internal consistency and SNR) in middle to late adolescence.

As such, these data contribute to an emerging body of work identifying developmentally-appropriate experimental paradigms that can be used to reliably capture the attention of and engage, thereby reliably elicit ERP measures in, adolescents. The Doors task, given its simplicity, has allowed it to be feasibly used in clinical populations (e.g., Bunford et al., 2021; Horan et al., 2012) and children (e.g., Kujawa et al., 2014; Schneider et al., 2016) while remaining a potent probe in adults (e.g., Weinberg et al., 2014). Our earlier results, obtained with the same sample as assessed here, are evidence of convergent validity (but also indicate distinction) between MID and Doors ERPs and self-reported reinforcement sensitivity and of incremental validity of the ERPs and self- reported reinforcement sensitivity in predicting these affective outcomes (Zubovics et al., 2021). Our current results suggest evidence of acceptable data quality, stability, and internal consistency of such ERPs, across custom and standardized preprocessing approaches. Together, others' and our own (Zubovics et al., 2021) prior and current findings inform a broader question of psychometrics; across studies, results suggest both the MID and the Doors task can be feasibly and reliably used with middle-late adolescents and that these reliable estimates appear to probe characteristics they are intended to probe, with the ERP measures evidently a valuable addition to a clinically-informative and multi-method measurement framework.

As noted, there is debate as to whether the MID task is appropriate for probing, beyond reward anticipation, reward response. Others have reported on decipherability of these two aspects of reward processing in both fMRI and ERP versions of the task (Knutson, Fong, Adams, et al., 2001; Novak et al., 2016; Novak & Foti, 2015). Capitalizing on our within-subject design, in exploratory analyses, we compared MID and Doors RewP values to determine the extent to which the two tasks probe the same phenomenon and which task elicits a greater neural response. Given the magnitude of obtained correlation coefficients, MID and Doors RewP to win ($r = 0.604$) and lose ($r = 0.726$) appear to be measuring strongly related but not identical phenomena (depending on isomorphism threshold). Further, MID appears to elicit a stronger RewP to both win and lose. As such, MID and Doors seem to probe almost the same if not the same neural responses to reward receipt and this response is stronger as probed by the MID task, likely as a result of greater engagement with the task, either due to more virtual money being at stake and/or less boredom. Due to poor internal consistency, results with the ΔRewP (low between-task correlation and no within-person difference) cannot be interpreted.

Our findings are informative for best practices and standards for ERP research with middle-late adolescence. Although results will need replication, exploration of explanatory hypotheses regarding cross-study (child/adult vs. adolescent) differences is warranted. Specifically, findings indicate that in middle-late adolescents, a greater number of trials is necessary to achieve stable and internally consistent estimates with acceptable SNR than recommended for children and adults. In adolescence, normative reward processing patterns are less consistent or stable than they are in childhood or adulthood. This likely corresponds to larger within-group differences in adolescents than in children or adults and this greater heterogeneity, in turn, necessitates a larger sample or in case of ERP, a greater number of trials, for group-level indices to reach comparable values. Although in cross-sectional studies, age did not moderate internal consistency of the RewP in children and young-middle adolescent (8–14 year-olds) girls (Luking et al., 2017), or across a very broad age-range (10–55 year-olds) (Ethridge & Weinberg, 2018), longitudinal research would be more informative about developmental differences in within-group heterogeneity. No such research has been conducted with the MID task and only one with the Doors task (Kujawa et al., 2018). In that research, youth exhibited greater variability in RewP to win during both their middle ($M = 17.27$, $SD = 10.00$) and their late ($M = 18.44$, $SD = 10.88$) adolescence, relative to their late childhood ($M = 6.48$, $SD = 7.42$) and this was also true for RewP to loss (middle $M = 12.33$, $SD = 9.09$ and late $M = 12.61$, $SD = 8.37$ adolescence vs. childhood $M = 1.26$, $SD = 6.39$)). As another explanatory hypothesis, in the current (and Zubovics et al., 2021) study, "money" earned in the tasks was exchanged for snacks whereas in other MID/Doors studies, money earned was given to participants as a "bonus". This may have impacted engagement and general ERP amplitude/psychometrics and may incidentally also explain why delta RewP values observed in the current study is smaller than is reported typically.

Others and our conclusions regarding necessary and sufficient number of trials is based on calculated data quality (SNR), stability, and/or internal consistency of examined ERP components. There are two issues with this approach and both indicate that determining the needed number of trials is not as simple as it may appear.

First, alpha is often employed as the end-all-be-all of reliability, but, as also evident from the data obtained here, a greater number of trials is needed to attain adequate SNR compared with the number of trials needed for adequate internal consistency.

Second, the advantages of and compromises across these approaches are important to consider. SNR is an index of the strength of a signal of interest relative to noise (Teplan, 2002). As the number of trials increases, SNR also

increases (Thigpen et al., 2017). SNR characterizes data quality and, as such, carries information about the precision of the ERP scores obtained (Clayson et al., 2021) but does not carry information about the extent to which that precision is sufficient for specific purposes, such as comparing ERPs across conditions, groups, or participants. Internal consistency is an index of the extent to which subunits of a measure (e.g., items or trials) correlate with each other, i.e., measure the same latent phenomenon. To a certain point, as the number of trials increases, internal consistency also increases. However, beyond that point, other factors may exert an effect (e.g., boredom, fatigue) and lead to the subunits no longer measuring the same latent phenomenon (e.g., ERPs being driven by reward responsiveness vs. by a combination of reward responsiveness and sustained attention), thereby resulting in a decrease in internal consistency. As such, *in aiming to achieve acceptable SNR, the number of trials should not be increased to an extent where acceptable internal consistency is sacrificed*.

Internal consistency thus carries information about the amount of true score (as opposed to error variance) that is contained in a measure that is available for association with outcomes (Levinson et al., 2017). Internal consistency can be assessed at the level of the group and of the individual; the former, group-level internal consistency, is a measure of between-person differences relative to the precision of scores for the group, and the latter, individual-level internal consistency, is a measure of the precision of a score for a person relative to between-person differences for a group (Clayson et al., 2021). As such, estimating group-level internal consistency will yield a single reliability estimate for the entire group, potentially masking low reliability for certain individuals (Clayson et al., 2021). Indeed, although it had been assumed that meaningful variability in ERPs are present, primarily, between rather than within persons, recent findings indicate that ERPs may change over the course of an experimental paradigm (Berry et al., 2019; Brush et al., 2018; Volpert-Esmond et al., 2018). Accordingly, as others have most recently recommended, *in research focused on assessing individual differences, such as in clinical or cognitive neuroscience studies, reliability is to be estimated at the level of the individual*, to help determine whether individual-level data are sufficiently reliable for valid statistical inference (Clayson et al., 2021).

Third, this approach is not informative about the number of trials necessary to detect between-condition or between-group differences with standard statistical analyses (Gehring et al., 2012), that is, *statistical power*. Data show that the number of trials recommended based calculated stability may be insufficient for detecting between-group differences (Fischer et al., 2017). As the aim of many (if not the majority of) ERP studies is to determine whether there are ERP differences across conditions, individuals, or groups, a critical issue beyond data quality, stability and reliability is determining statistical power – the effect size and number of participants and trials needed to detect between-conditions or between-groups effects. In the field of neuroimaging, even less data are available on statistical power than on reliability (or validity). As there is a tendency for neuroscience studies to be underpowered, there is decreased likelihood that existing effects are detected but an increased likelihood that detected effects are overestimated (Button et al., 2013; Groppe, 2017). As such, both individual-level assessment of psychometrics and estimation of power are a foremost problem to address in this line of research.

5 | CONCLUSION

Both MID and Doors ERPs achieve acceptable data quality, stability, and internal consistency in less than 50 trials in middle-late adolescents. Although custom preprocessing pipelines are inherently subjective, data having been largely replicated using a standardized script increases confidence in generalizability. In combination with earlier findings indicating convergent and incremental validity, there is emerging evidence that these tasks generate psychometrically sound measures of reward anticipation and response. MID and Doors appear to probe the same, but MID appears to probe a stronger, reward response. Next steps in this line of work will be to include measures of individual-level psychometrics and statistical power.

ACKNOWLEDGMENTS

This research was funded by an MTA Lendület ("Momentum") Grant awarded to NB (#LP2018-3/2018). During the preparation of this article, AR was supported by the New National Excellence Program 2021/2022 (ÚNKP-21-3-II-SE-53) and JMR was supported by the Higher Education Institutional Excellence Program of the Ministry of Human Capacities in Hungary, within the framework of the Neurology thematic program of Semmelweis University. This research was further supported by the research data archiving pilot grant from the Library and Information Centre of the Hungarian Academy of Sciences, the Hungarian Research Data Alliance and the Eötvös Loránd Research Network.

AUTHOR CONTRIBUTIONS

György Hámori: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; software; visualization; writing – original draft. **Alexandra Rádosi:** Investigation; methodology;

project administration; writing – review and editing. **Bea Pászthy:** Supervision; writing – review and editing. **János M Réthelyi:** Supervision; writing – review and editing. **István Ulbert:** Supervision; writing – review and editing. **Richárd Fiáth:** Conceptualization; data curation; formal analysis; software; supervision; writing – original draft. **Nóra Bunford:** Conceptualization; data curation; funding acquisition; project administration; resources; supervision; writing – original draft.

DATA AVAILABILITY STATEMENT

Datasets and codes generated and/or analyzed for the current study are available from the corresponding author on reasonable request. We confirm that for all experiments, we reported all conditions, measures, and data exclusions.

ORCID

Nóra Bunford https://orcid.org/0000-0002-9645-604X

REFERENCES

Auerbach, R. P., Bondy, E., Stanton, C. H., Webb, C. A., Shankman, S. A., & Pizzagalli, D. A. (2016). Self-referential processing in adolescents: Stability of behavioral and ERP markers. *Psychophysiology*, *53*, 1398–1406. https://doi.org/10.1111/psyp.12686

Ball, T. M. M., Stein, M. B. B., & Paulus, M. P. P. (2014). Toward the application of functional neuroimaging to individualized treatment for anxiety and depression. *Depression and Anxiety*, *31*, 920–933. https://doi.org/10.1002/da.22299

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Berry, M. P., Tanovic, E., Joormann, J., & Sanislow, C. A. (2019). Relation of depression symptoms to sustained reward and loss sensitivity. *Psychophysiology*, *56*, e13364. https://doi.org/10.1111/psyp.13364

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K. M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*(16), 1–20. https://doi.org/10.3389/fninf.2015.00016

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*, 1–16. https://doi.org/10.1111/psyp.13049

Bress, J. N., Foti, D., Kotov, R., Klein, D. N., & Hajcak, G. (2013). Blunted neural response to rewards prospectively predicts depression in adolescent girls. *Psychophysiology*, *50*, 74–81. https://doi.org/10.1111/j.1469-8986.2012.01485.x

Bress, J. N., Meyer, A., & Proudfit, G. H. (2015). The stability of the feedback negativity and its relationship with depression during childhood and adolescence. *Development and Psychopathology*, *27*, 1285–1294. https://doi.org/10.1017/s0954579414001400

Bress, J. N., Smith, E., Foti, D., Klein, D., & Hajcak, G. (2012). Neural response to reward and depressive symptoms in late childhood to early adolescence. *Biological Psychology*, *89*, 156–162. https://doi.org/10.1016/j.biopsycho.2011.10.004.Neural

Broyd, S. J., Richards, H. J., Helps, S. K., Chronaki, G., Bamford, S., & Sonuga-Barke, E. J. S. (2012). Electrophysiological markers of the motivational salience of delay imposition and escape. *Neuropsychologia*, *50*, 965–972. https://doi.org/10.1016/j.neuropsychologia.2012.02.003

Brush, C. J., Ehmann, P. J., Hajcak, G., Selby, E. A., & Alderman, B. L. (2018). Using multilevel modeling to examine blunted neural responses to reward in major depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(12), 1032–1039. https://doi.org/10.1016/j.bpsc.2018.04.003

Bunford, N., Kujawa, A., Dyson, M., Olino, T., & Klein, D. N. (2021). Developmental pathways from preschool temperament to early adolescent ADHD symptoms through initial responsiveness to reward. *Development and Psychopathology*, *16*, 1–13. https://doi.org/10.1017/S0954579420002199

Bunford, N., Kujawa, A., Fitzgerald, K. D., Swain, J. E., Hanna, G. L., Koschmann, E., Simpson, D., Connolly, S., Monk, C. S., & Phan, K. L. (2017). Neural reactivity to angry faces predicts treatment response in pediatric anxiety. *Journal of Abnormal Child Psychology*, *45*, 385–395. https://doi.org/10.1007/s10802-016-0168-2

Bunford, N., Kujawa, A., Swain, J. E., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. L. (2017). Attenuated neural reactivity to happy faces is associated with rule breaking and social problems in anxious youth. *European Child & Adolescent Psychiatry*, *26*, 215–230. https://doi.org/10.1007/s00787-016-0883-9

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, 365–376. https://doi.org/10.1038/nrn3475

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*, *49*, 659–664. https://doi.org/10.1111/j.1469-8986.2011.01349.x

Cauffman, E., Shulman, E. P., Steinberg, L., Claus, E., Banich, M. T., Graham, S., & Woolard, J. (2010). Age differences in affective decision making as indexed by performance on the Iowa gambling task. *Developmental Psychology*, *46*, 193–207. https://doi.org/10.1037/a0016128

Chronaki, G., Soltesz, F., Benikos, N., & Sonuga-Barke, E. J. S. (2017). An electrophysiological investigation of reinforcement effects in attention deficit/hyperactivity disorder: Dissociating cue sensitivity from down-stream effects on target engagement and performance. *Developmental Cognitive Neuroscience*, *28*, 12–20. https://doi.org/10.1016/j.dcn.2017.10.003

Clayson, P. E., Brush, C. J., & Hajcak, G. (2021). Data quality and reliability metrics for event-related potentials (ERPs): The utility of subject-level reliability. *International Journal of Psychophysiology*, *165*, 121–136. https://doi.org/10.1016/j.ijpsycho.2021.04.004

Cowley, B. U., Korpela, J., & Torniainen, J. (2017). Computational testing for automated preprocessing: A matlab toolbox to enable large scale electroencephalography data processing. *PeerJ Computer Science*, *3*, e108. https://doi.org/10.7717/peerj-cs.108

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Crowley, M. J., Wu, J., Hommer, R. E., South, M., Molfese, P. J., Fearon, R. M. P., & Mayes, L. C. (2013). A developmental

study of the feedback-related negativity from 10–17 years: Age and sex effects for reward versus non-reward. *Developmental Neuropsychology*, *38*, 595–612. https://doi.org/10.1080/87565641.2012.694512

da Cruz, J. R., Chicherov, V., Herzog, M. H., & Figueiredo, P. (2018). An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clinical Neurophysiology*, *129*, 1427–1437. https://doi.org/10.1016/j.clinph.2018.04.600

De Los Reyes, A. (2017). Inaugural editorial: Making the journal of clinical child & adolescent psychology your "Home Journal". *Journal of Clinical Child and Adolescent Psychology*, *46*(1), 1–10. https://doi.org/10.1080/15374416.2016.1266649

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*, e13580. https://doi.org/10.1111/psyp.13580

Doehrmann, O., Ghosh, S. S., Polli, F. E., Reynolds, G. O., Horn, F., Keshavan, A., Triantafyllou, C., Saygin, Z. M., Whitfield-Gabrieli, S., Hofmann, S. G., Pollack, M., & Gabrieli, J. D. (2013). Predicting treatment response in social anxiety disorder from functional magnetic resonance imaging. *JAMA Psychiatry*, *70*, 87–97. https://doi.org/10.1001/2013.jamapsychiatry.5

Dunning, J. P., & Hajcak, G. (2007). Error-related negativities elicited by monetary loss and cues that predict loss. *Neuroreport*, *18*, 1875–1878. https://doi.org/10.1097/WNR.0b013e3282f0d50b

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51–100. https://doi.org/10.1006/obhd.1994.1029

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., Ireland, D., Ramrakha, S., Poulton, R., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage*, *189*, 516–532. https://doi.org/10.1016/j.neuroimage.2019.01.068

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, *31*(7), 792–806. https://doi.org/10.1177/0956797620916786

Ernst, M. (2014). The triadic model perspective for the study of adolescent motivated behavior. *Brain and Cognition*, *89*, 104–111. https://doi.org/10.1016/j.bandc.2014.01.006

Ernst, M., & Spear, L. P. (2009). Reward systems. In M. de Haan & M. Gunnar (Eds.), *Handbook of developmental social neuroscience* (pp. 324–341). Guilford Press.

Ethridge, P., Kujawa, A., Dirks, M. A., Arfer, K. B., Kessel, E. M., Klein, D. N., & Weinberg, A. (2017). Neural responses to social and monetary reward in early adolescence and emerging adulthood. *Psychophysiology*, *54*, 1786–1799. https://doi.org/10.1111/psyp.12957

Ethridge, P., & Weinberg, A. (2018). Psychometric properties of neural responses to monetary and social rewards across development. *International Journal of Psychophysiology*, *132*, 311–322. https://doi.org/10.1016/j.ijpsycho.2018.01.011

Fischer, A. G., Klein, T. A., & Ullsperger, M. (2017). Comparing the error-related negativity across groups: The impact of error- and trial-number differences. *Psychophysiology*, *54*, 998–1009. https://doi.org/10.1111/psyp.12863

Foti, D., & Hajcak, G. (2012). Genetic variation in dopamine moderates neural response during reward anticipation and delivery: Evidence from event-related potentials. *Psychophysiology*, *49*, 617–626. https://doi.org/10.1111/j.1469-8986.2011.01343.x

Foti, D., & Hajcak, G. (2009). Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. *Biological Psychology*, *81*, 1–8. https://doi.org/10.1016/j.biopsycho.2008.12.004

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Response to commentary. *Human Brain Mapping*, *32*, 2267–2269. https://doi.org/10.1002/hbm.21357

Frewen, P. A., Dozois, D. J. A., & Lanius, R. A. (2008). Neuroimaging studies of psychological interventions for mood and anxiety disorders: Empirical and methodological review. *Clinical Psychology Review*, *28*, 228–246. https://doi.org/10.1016/j.cpr.2007.05.002

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction* (2nd ed.). Sage Publications, Inc.

Gabard-Durnam, L. J., Leal, A. S. M., Wilkinson, C. L., & Levin, A. R. (2018). The Harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, *12*(97), 1–24. https://doi.org/10.3389/fnins.2018.00097

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*, 11–26. https://doi.org/10.1016/j.neuron.2014.10.047

Galván, A. (2013). The teenage brain: Sensitivity to rewards. *Current Directions in Psychological Science*, *22*, 88–93. https://doi.org/10.1177/0963721413480859

Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN/ne). In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 231–291). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0120

Groom, M. J., Scerif, G., Liddle, P. F., Batty, M. J., Liddle, E. B., Roberts, K. L., Cahill, J. D., Liotti, M., & Hollis, C. (2010). Effects of motivation and medication on electrophysiological markers of response inhibition in children with attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *67*, 624–631. https://doi.org/10.1016/j.biopsych.2009.09.029

Groppe, D. M. (2017). Combating the scientific decline effect with confidence (intervals). *Psychophysiology*, *54*(1), 139–145. https://doi.org/10.1111/psyp.12616

Hajcak, G., Weinberg, A., Macnamara, A., & Foti, D. (2011). ERPs and the study of emotion.Pdf. In S. J. Luck & E. S. Kappenman (Eds.), *Handbook of event-related potential components* (pp. 441–472). Oxford University Press.

Hämmerer, D., Li, S. C., Müller, V., & Lindenberger, U. (2011). Life span differences in electrophysiological correlates of monitoring gains and losses during probabilistic reinforcement learning. *Journal of Cognitive Neuroscience*, *23*, 579–592. https://doi.org/10.1162/jocn.2010.21475

Horan, W. P., Foti, D., Hajcak, G., Wynn, J. K., & Green, M. F. (2012). Impaired neural response to internal but not external

feedback in schizophrenia. *Psychological Medicine*, *42*, 1637–1647. https://doi.org/10.1017/S0033291711002819

Kessel, E. M., Kujawa, A., Hajcak Proudfit, G., & Klein, D. N. (2015). Neural reactivity to monetary rewards and losses differentiates social from generalized anxiety in children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*, 792–800. https://doi.org/10.1111/jcpp.12355

Klawohn, J., Meyer, A., Weinberg, A., & Hajcak, G. (2020). Methodological choices in event-related potential (ERP) research and their impact on internal consistency reliability and individual differences: An examination of the error-related negativity (ERN) and anxiety. *Journal of Abnormal Psychology*, *129*, 29–37. https://doi.org/10.1037/abn0000458

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport*, *12*, 3683–3687. https://doi.org/10.1097/00001756-200112040-00016

Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *NeuroImage*, *18*, 263–272. https://doi.org/10.1016/S1053-8119(02)00057-5

Knutson, B., Fong, G. W., Hommer, D., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *The Journal of Neuroscience*, *21*, RC159. https://doi.org/10.1523/JNEUROSCI.21-16-j0002.2001 [pii].

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews. Neuroscience*, *6*, 691–702. https://doi.org/10.1038/nrn1747

Kujawa, A., Carroll, A., Mumper, E., Mukherjee, D., Kessel, E. M., Olino, T., Hajcak, G., & Klein, D. N. (2018). A longitudinal examination of event-related potentials sensitive to monetary reward and loss feedback from late childhood to middle adolescence. *International Journal of Psychophysiology*, *132*, 323–330. https://doi.org/10.1016/j.ijpsycho.2017.11.001

Kujawa, A., Hajcak, G., & Klein, D. N. (2019). Reduced reward responsiveness moderates the effect of maternal depression on depressive symptoms in offspring: Evidence across levels of analysis. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *60*, 82–90. https://doi.org/10.1111/jcpp.12944

Kujawa, A., Proudfit, G. H., & Klein, D. N. (2014). Neural reactivity to rewards and losses in offspring of mothers and fathers with histories of depressive and anxiety disorders. *Journal of Abnormal Psychology*, *123*, 287–297. https://doi.org/10.1037/a0036285

Kujawa, A., Proudfit, G. H. H., Hajcak, G., Kessel, E. M. M., Dyson, M., Olino, T., & Klein, D. N. N. (2015). Neural reactivity to monetary rewards and losses in childhood: Longitudinal and concurrent associations with observed and self-reported positive emotionality. *Biological Psychology*, *104*, 41–47. https://doi.org/10.1016/j.biopsycho.2014.11.008

Kujawa, A., Proudfit, G. H. H., Kessel, E. M. M., Dyson, M., Olino, T., & Klein, D. N. N. (2015). Neural reactivity to monetary rewards and losses in childhood: Longitudinal and concurrent associations with observed and self-reported positive emotionality. *Biological Psychology*, *104*, 41–47. https://doi.org/10.1016/j.biopsycho.2014.11.008

Kujawa, A., Smith, E., Luhmann, C., & Hajcak, G. (2013). The feedback negativity reflects favorable compared to nonfavorable outcomes based on global, not local, alternatives. *Psychophysiology*, *50*, 134–138. https://doi.org/10.1111/psyp.12002

Kujawa, A., Weinberg, A., Bunford, N., Fitzgerald, K. D., Hanna, G. L., Monk, C. S. S., Kennedy, A. E., Klumpp, H., Hajcak, G., Phan, K. L. L., Swain, J. E., Monk, C. S., Hajcak, G., & Phan, K. L. (2016). Error-related brain activity in youth and young adults with generalized or social anxiety disorder before and after treatment. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *71*, 162–168. https://doi.org/10.1016/j.pnpbp.2016.07.010

Kujawa, A., Weinberg, A., Hajcak, G., & Klein, D. N. (2013). Differentiating event-related potential components sensitive to emotion in middle childhood: Evidence from temporal-spatial PCA. *Developmental Psychobiology*, *55*, 539–550. https://doi.org/10.1002/dev.21058

Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology*, *41*, 1–14. https://doi.org/10.1007/s10802-012-9659-y

Layne, A. E., Bernstein, G. A., Egan, E. A., & Kushner, M. G. (2003). Predictors of treatment response in anxious-depressed adolescents with school refusal. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 319–326. https://doi.org/10.1097/00004583-200303000-00012

Levin, A. R., Leal, A. S. M., Gabard-Durnam, L. J., & O'Leary, H. M. (2018). BEAPP: The batch electroencephalography automated processing platform. *Frontiers in Neuroscience*, *12*(513), 1–13. https://doi.org/10.3389/fnins.2018.00513

Levinson, A. R., Speed, B. C., Infantolino, Z. P., & Hajcak, G. (2017). Reliability of the electrocortical response to gains and losses in the doors task. *Psychophysiology*, *54*, 601–607. https://doi.org/10.1111/psyp.12813

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). University of Wisconsin Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (II. ed.). The MIT Press.

Lukie, C. N., Montazer-Hojat, S., & Holroyd, C. B. (2014). Developmental changes in the reward positivity: An electrophysiological trajectory of reward processing. *Developmental Cognitive Neuroscience*, *9*, 191–199. https://doi.org/10.1016/j.dcn.2014.04.003

Luking, K. R., Gilbert, K., Kelly, D., Kappenman, E. S., Hajcak, G., Luby, J. L., & Barch, D. M. (2021). The relationship between depression symptoms and adolescent neural response during reward anticipation and outcome depends on developmental timing: Evidence from a longitudinal study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(5), 527–535. https://doi.org/10.1016/j.bpsc.2020.11.001

Luking, K. R., Nelson, B. D., Infantolino, Z. P., Sauder, C. L., & Hajcak, G. (2017). Internal consistency of functional magnetic resonance imaging and electroencephalography measures of reward in late childhood and early adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *2*(3), 289–297. https://doi.org/10.1016/j.bpsc.2016.12.004

Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., & Rodriguez-Fornells, A. (2011). On the number of trials needed for a

stable feedback-related negativity. *Psychophysiology*, *48*, 852–860. https://doi.org/10.1111/j.1469-8986.2010.01152.x

Morgan, J. K., Olino, T. M., McMakin, D. L., Ryan, N. D., & Forbes, E. E. (2013). Neural response to reward as a predictor of increases in depressive symptoms in adolescence. *Neurobiology of Disease*, *52*, 66–74. https://doi.org/10.1016/j.nbd.2012.03.039

National Institute of Mental Health. National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (RDoC), National Institute of Mental Health, 2016. Behavioral Assessment Methods for RDoC Constructs.

Novak, B. K., Novak, K. D., Lynam, D. R., & Foti, D. (2016). Individual differences in the time course of reward processing: Stage-specific links with depression and impulsivity. *Biological Psychology*, *119*, 79–90. https://doi.org/10.1016/j.biopsycho.2016.07.008

Novak, K. D., & Foti, D. (2015). Teasing apart the anticipatory and consummatory processing of monetary incentives: An event-related potential study of reward dynamics. *Psychophysiology*, *52*, 1470–1482. https://doi.org/10.1111/psyp.12504

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. T. (1967). *Psychometric theory*. McGraw-Hill.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Oumeziane, B. A., Jones, O., & Foti, D. (2019). Neural sensitivity to social and monetary reward in depression: Clarifying general and domain-specific deficits. *Frontiers in Behavioral Neuroscience*, *13*(199), 1–16. https://doi.org/10.3389/fnbeh.2019.00199

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046

Rodrigues, J., Weiß, M., Hewig, J., & Allen, J. J. B. (2021). EPOS: EEG processing open-source scripts. *Frontiers in Neuroscience*, *15*, 660449. https://doi.org/10.3389/fnins.2021.660449

Schneider, S. K., Buka, S. L., Dash, K., Winickoff, J. P., & Donnell, L. O. (2016). Community reductions in youth smoking after raising the minimum tobacco sales age to 21. *Tobacco Control*, *25*, 355–359. https://doi.org/10.1136/tobaccocontrol-2014-052207

Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantoniou, D. K., & Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, *47*, 260–270. https://doi.org/10.1111/j.1469-8986.2009.00942.x

Shulman, E. P., Smith, A. R., Silva, K., Icenogle, G., Duell, N., Chein, J., & Steinberg, L. (2016). The dual systems model: Review, reappraisal, and reaffirmation. *Developmental Cognitive Neuroscience*, *17*, 103–117.

Siegle, G. J. (2011). Beyond depression commentary: Wherefore art thou, depression clinic of tomorrow? *Clinical Psychology: Science and Practice*, *18*, 305–310. https://doi.org/10.1111/j.1468-2850.2011.01261.x

Silverman, M. H., Jedd, K., & Luciana, M. (2015). Neural networks involved in adolescent reward processing: An activation likelihood estimation meta-analysis of functional neuroimaging studies. *NeuroImage*, *122*, 427–439. https://doi.org/10.1016/j.neuroimage.2015.07.083

Spear, L. (2013). Adolescent neurodevelopment. In *Adolescent neurodevelopment*. J. Adolesc. https://doi.org/10.1016/j.jadohealth.2012.05.006

Spear, L. P. (2018). Effects of adolescent alcohol consumption on the brain and behaviour. *Nature Reviews. Neuroscience*, *19*, 197–214. https://doi.org/10.1038/nrn.2018.10

Stice, E., Yokum, S., & Burger, K. S. (2013). Elevated reward region responsivity predicts future substance use onset but not overweight/obesity onset. *Biological Psychiatry*, *73*, 869–876. https://doi.org/10.1016/j.biopsych.2012.11.019

Teplan, M. (2002). Fundamentals of EEG measurement. *Measurement Science Review*, *2*, 1. https://doi.org/10.1016/j.bpsc.2021.10.017

Thigpen, N. N., Kappenman, E. S., & Keil, A. (2017). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology*, *54*(1), 123–138. https://doi.org/10.1111/psyp.12629

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, *55*, e13044. https://doi.org/10.1111/psyp.13044

Weinberg, A., Riesel, A., & Proudfit, G. H. (2014). Show me the money: The impact of actual rewards and losses on the feedback negativity. *Brain and Cognition*, *87*, 134–139. https://doi.org/10.1016/j.bandc.2014.03.015

Whitfield-Gabrieli, S., Ghosh, S. S., Nieto-Castanon, A., Saygin, Z., Doehrmann, O., Chai, X. J., Reynolds, G. O., Hofmann, S. G., Pollack, M. H., & Gabrieli, J. D. E. (2015). Brain connectomics predict response to treatment in social anxiety disorder. *Molecular Psychiatry*, *1–6*, 680–685. https://doi.org/10.1038/mp.2015.109

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. EMBS. https://doi.org/10.1109/EMBC.2015.7319296

Zubovics, E. A., Fiáth, R., Rádosi, A., Pászthy, B., Réthelyi, J. M., Ulbert, I., & Bunford, N. (2021). Neural and self-reported reward responsiveness are associated with dispositional affectivity and emotion dysregulation in adolescents with evidence for convergent and incremental validity. *Psychophysiology*, *58*, e13723. https://doi.org/10.1111/psyp.13723

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1 MID Cue P3 across task orders, with MID first (left column) and Doors first (right column)

Figure S2 MID Target P3 across task orders, with MID first (left column) and Doors first (right column)

Figure S3 MID SPN across task orders, with MID first (left column) and Doors first (right column)

Figure S4 MID RewP across task orders, with MID first (left column) and Doors first (right column)

Figure S5 Doors RewP across task orders, with MID first (left column) and Doors first (right column)
Figure S6 Doors RewP with $n = 41$ participants
Figure S7 MID Cue P3 with the custom (left column) and MADE (right column)
Figure S8 MID Target P3 with the custom (left column) and MADE (right column) preprocessing pipelines
Figure S9 MID SPN with the custom (left column) and MADE (right column) preprocessing pipelines
Figure S10 MID RewP with the custom (left column) and MADE (right column) preprocessing pipelines
Figure S11 Doors RewP with the custom (left column) and MADE (right column) preprocessing pipelines
Table S1 Descriptive statistics (in μVs) for all ERP components across the first 74 and across all trials using the MADE pipeline
Table S2 Descriptive statistics (in μVs) for all ERP components across the first 74 and across all trials using the MADE pipeline

How to cite this article: Hámori, G., Rádosi, A., Pászthy, B., Réthelyi, J. M., Ulbert, I., Fiáth, R. & Bunford, N. (2022). Reliability of reward ERPs in middle-late adolescents using a custom and a standardized preprocessing pipeline. *Psychophysiology*, 59, e14043. https://doi.org/10.1111/psyp.14043