Aus dem Fachbereich Medizin
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

betreut am
Zentrum der Psychischen Gesundheit
Klinik für Psychiatrie, Psychosomatik und Psychotherapie
Direktor: Prof. Dr. Andreas Reif

Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task.

Dissertation
zur Erlangung des Doktorgrades der Medizin
des Fachbereichs Medizin
der Johann Wolfgang Goethe-Universität
Frankfurt am Main

vorgelegt von
Franziska Suchan

aus Hachenburg

Frankfurt am Main, 2020

Dekan/in:	Herr Prof. Dr. Stefan Zeuzem
Referent/in:	Herr Prof. Dr. Michael Plichta
Korreferent/in:	
ggf. 2. Korreferent/in:	
Tag der mündlichen Prüfung:	

Danksagung:

An erster Stelle möchte ich meinen Eltern danken, die mir ein sorgenfreies Studium ermöglicht haben, auf die ich mich in jeder Lebenssituation verlassen kann und die mir immer mit Rat und Tat zur Seite stehen.

Außerdem danke ich Prof. Plichta, der es verstand, mich zu begeistern, zu fordern und für jedes Problem eine Lösung parat hatte.

Ein großes Dankeschön geht an meine Freunde von Goethes KOMMchester, die mein Studium zu einer wunderbaren Zeit voller Musik und Freundschaft gemacht haben.

Inhaltsverzeichnis

1 Table index

2 Figure index

3 Zusammenfassung (Deutsch)

Aktuelle Studien zeigen, dass die *error-related negativity* (ERN), ein ereigniskorreliertes Potenzial, welches nach fehlerhaften Reaktionen im EEG messbar ist, bei verschiedenen psychiatrischen Störungen verändert ist und dazu beitragen kann, Behandlungsentscheidungen zu unterstützen. Damit ist die ERN ein vielversprechender Kandidat für einen psychiatrischen Biomarker. Grundlegende methodische Anforderungen an einen Biomarker sind standardisierte und reliable Messungen. Zusätzliche Psychiatrie-spezifische Anforderungen sind die zeiteffiziente und patientenfreundliche Messung.

Ziel dieser Studie ist es, die ERN reliabel, zeiteffizient und patientenfreundlich zu messen, um ihren Einsatz im klinischen Alltag zu etablieren.

Gesunde Probanden (*N*=27) haben eine modifizierte *Eriksen Flanker Task* mit adaptivem Reaktionszeitfenster und ausschließlich inkongruenten Stimuli bearbeitet, welche die Fehleranzahl maximiert. Alle Probanden wurden mit dem *Mini International Neuropsychiatric Interview* (M.I.N.I.) auf psychische Gesundheit untersucht. Die ersten *N*=12 Probanden waren Teil einer Pilotstudie, weitere *N*=14 Probanden wurden in die Auswertung der Hauptanalyse eingeschlossen (ein Proband wurde wegen technischer Probleme ausgeschlossen). In einem Test-Retest-Design mit zwei Messungen im Abstand von 28 Tagen wurde die Reliabilität der ERN gemessen. Die externe Validität wurde durch Replikation von aus der Literatur bekannten Korrelationen der ERN-Amplitude mit (1) der Fehleranzahl und (2) negativem Affekt überprüft. Dazu wurde der emotionale Zustand des Probanden zum Zeitpunkt der Messung anhand des *Positive and Negative Affect Schedule* erfasst. Um die klinische Praktikabilität der Aufgabe zu optimieren, wurde

untersucht, inwieweit die Aufgabe gekürzt werden kann, ohne eine minimale Reliabilität der ERN von 0.80 zu unterschreiten. Eine exzellente Reliabilität der ERN (Interkorrelationskoeffizient=0.806-0.947) konnte bestätigt und spezifische Korrelationsmuster repliziert werden (ERN Amplitude mit der relativen Fehleranzahl: r=0.394; p=0.082. ERN Amplitude mit dem negativen Affekt: r=-0.583, p=0.014). Die Aufgabe kann auf eine patientenfreundliche und im klinischen Alltag praktikable Länge von nur 8 Minuten gekürzt werden, wobei die Reliabilität einen Wert von 0.80 nicht unterschreitet.

Zusammenfassend lässt sich sagen, dass die modifizierte *Eriksen Flanker Task* eine reliable und effiziente Messung der ERN und die Etablierung als psychiatrischen Biomarker ermöglicht.

4 Abstract (English)

There is accumulating evidence that the *error-related negativity* (ERN), an event-related potential elicited after erroneous actions, is altered in different psychiatric disorders and may help to guide treatment options. Thus, the ERN is a promising candidate as a psychiatric biomarker. Basic methodological requirements for a biomarker are standardized and reliable measurements. Additional psychiatry specific requirements are time efficiency and patient-friendliness.

The aim of the present study is to establish ERN acquisition in a reliable, time-efficient and patient-friendly way for use in clinical practice.

Healthy subjects (*N*=27) performed a modified *Eriksen Flanker Task* with adaptive reaction time window and only incongruent stimuli that maximizes the number of errors. All participants were tested for mental health by the *Mini International Neuropsychiatric Interview* (M.I.N.I.). The first *N*=12 subjects were part of a pilot study and further *N*=14 subjects were included for analysis (one subject was excluded due to technical problems). In a test-retest design with two sessions separated by 28 days the reliability of the ERN has been assessed. To ensure external validity, we aimed to replicate previously reported correlation patterns of ERN amplitude with (1) number of errors and (2) negative affect. State affect of each subject was measured by the *Positive and Negative Affect Schedule.* In order to optimize the clinical use of the task, we determined to which extent the task can be shortened while keeping reliability >0.80.

We found excellent reliability of the ERN (intraclass correlation coefficient =0.806-0.947) and replicated specific correlation patterns (ERN amplitude with relative number of errors: r=0.394; p=0.082; ERN amplitude with negative affect: r=-0.583, p=0.014). The task can be shortened to a patient-

friendly and clinically feasible length of only 8 minutes keeping reliability >0.80.

To conclude, the present modified task provides reliable and efficient recording of the ERN, facilitating its use as a psychiatric biomarker.

5 Abkürzungsverzeichnis

ACC	Anterior cingulate cortex/ anteriorer cingulärer Cortex
ANOVA	Analysis of variance
CRN	Correct-related negativity
Cz	Midline central
DC	Direct-coupled
EEG	Electroencephalography/ Elektroenzephalogramm
EKP	Ereignis-korreliertes Potenzial
EOG	Electrooculogram
ERN	Error-related negativity
ERP	Event-related potential
FPz	Midline frontopolar
FRN	Feedback-related negativity
Fz	Midline frontal
ICC	Intraclass correlation coefficient/ Intraklassenkorrelationskoeffizient
M.I.N.I.	Mini International Neuropsychiatric Interview
PANAS	Positive and Negative Affect Schedule
PRO	Predicted-response outcome
RDoC	Research Domain Criteria
RT	Reaction time
SE	Standard error

6 Übergreifende Zusammenfassung

Das Elektroenzephalogramm (EEG) ist ein nichtinvasives, neurophysiologisches Verfahren, welches kortikale neuronale Aktivität mit hoher zeitlicher Auflösung misst. Dabei werden hirnelektrische Ströme als Potenzialdifferenz zwischen zwei Elektroden abgeleitet.[1]

Das EEG hat sowohl in der Forschung als auch in der klinischen Diagnostik einen hohen Stellenwert. Es ist breit verfügbar, kostengünstig und risikoarm. Zur neurologischen Epilepsiediagnostik und Schlafpolygraphie ist das EEG unverzichtbar. Aber auch in der Psychiatrie hat es nicht nur in der neurologischen Ausschlussdiagnostik Relevanz, sondern unterstützt bei der Differenzialdiagnose von organisch psychischen Störungen, wie Demenz, Delir oder dem nichtkonvulsiven Status epilepticus. Außerdem kann es in der Psychopharmakotherapie Hinweise auf neurotoxische Medikamenteneffekte oder Eigenmedikation geben.[1–3]

Neben der Frequenzanalyse von EEG-Daten können ereigniskorrelierte Potenziale (EKP) wichtige Informationen über kognitive Prozesse geben. Diese EKP werden durch sensorische, kognitive oder motorische Ereignisse ausgelöst und spiegeln die summierte Aktivität postsynaptischer Potenziale, von synchron feuernden, ähnlich orientierten, kortikalen Neuronen wieder[4]. Bei EEG-Messungen werden diese Potenziale repetitiv evoziert und im Anschluss die Mittelwerte analysiert.[1]

Aktuelle Forschung zeigt, dass besonders die *error-related negativity* (ERN) in der klinisch psychiatrischen Praxis zukünftig eine wichtige Rolle spielen könnte.

Die ERN ist ein negatives EKP, welches innerhalb von 100ms nach einer fehlerhaften Reaktion in frontozentralen Mittellinienelektroden gemessen werden kann[5,6] und

vermutlich im anterioren cingulären Cortex (ACC) entsteht[7–9] (Abbildung 1).

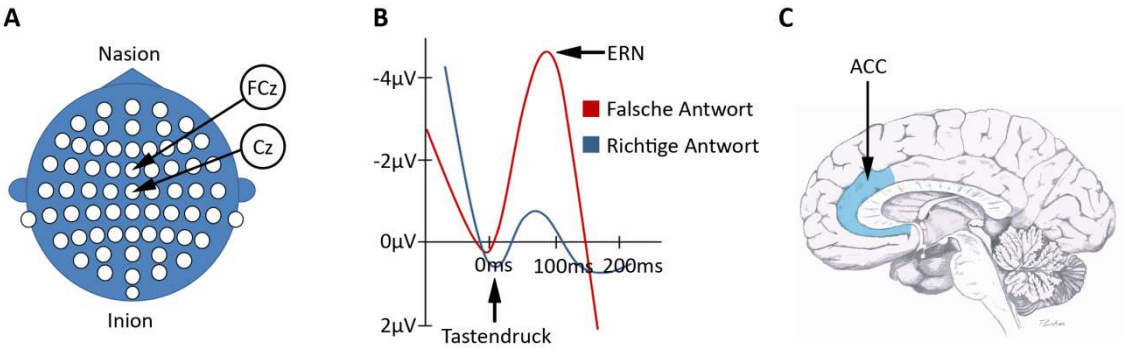


Abbildung 1 – (A) Position der EEG Elektroden. (B) Schematische Darstellung der *error-related negativity* (ERN). (C) Gehirn im Sagittalschnitt mit Markierung des anterioren cingulären Cortex (ACC) in blau.

Verschiedene Modelle versuchen die Funktion der ERN als ein Fehlererkennungssystem[10], einen Konflikterkennungsprozess[11] oder als Teil von operanter Konditionierung[12] zu beschreiben. Außerdem wurde kürzlich auf Grundlage des *forward models*[13] vermutet, dass die ERN Fehlerprognosen widerspiegeln könnte. Diese Vermutung steht im Einklang mit dem *predicted response-outcome* (PRO) Model[14], welches die ERN als eine Art Überraschungssignal interpretiert. Dieses wird ausgelöst, wenn ein Ereignis nicht mit der vorangegangenen Erwartungshaltung übereinstimmt.

Verschiedene Faktoren beeinflussen die ERN-Amplitude, dazu gehört die Leistung des einzelnen Probanden: Eine höhere Fehlerrate führt zu einer geringeren ERN-Amplitude[15,16]. Auch die Aufgabenstruktur ist relevant: a) Kongruente Stimuli führen zu einer höheren ERN-Amplitude als inkongruente Stimuli[17]; b) soll der Proband die Aufgabe eher präzise als schnell bearbeiten, führt dies zu einer höheren ERN-Amplitude[18].

Zudem haben negativer Affekt[19,20] und verschiedene psychiatrische Störungen[21–24] Auswirkung auf die ERN. Aktuelle Forschungen zeigen, dass die ERN a) Internalisierungsstörungen[25] wie z.B. Angststörung während der Adoleszenz[26,27] vorhersagt; b) ein Therapieansprechen nachweisen[21,22,28–30] und c) Behandlungsentscheidungen unterstützen kann[31]. Auf Grundlage dieser Studien könnte die ERN als neurophysiologischer Parameter in Zukunft prognostische, diagnostische und therapeutische Relevanz haben und den Stellenwert eines psychiatrischen Biomarkers einnehmen. Dies würde die Einsatzmöglichkeit des EEG von einem diagnostischen Instrument auf den Bereich der Prävention und Therapieentscheidung ausdehnen.
Eine Grundvoraussetzung für einen Biomarker ist dessen reliable Messung. Nur wenige Studien haben bis jetzt die Reliabilität der ERN unter Anwendung verschiedener Varianten der *Eriksen Flanker Task* untersucht und Intraklassen-Korrelationskoeffizienten (ICC) zwischen 0.62 und 0.74 gefunden[32–36].

Mit der vorliegenden Studie soll die Test-Retest-Reliabilität der ERN mit zwei Messungen im Abstand von 28 Tagen unter Anwendung einer modifizierten *Eriksen Flanker Task* mit adaptivem Reaktionszeitfenster[37,38] untersucht werden. Bei dieser Aufgabe muss der Proband einen zentralen Pfeil von umgebenden, irritierenden Pfeilen unterscheiden. Das adaptive Reaktionszeitfenster (siehe unten) der Task soll die Reliabilität aufgrund höherer Fehlerraten maximieren[36].
Um die Validität der ERN zu prüfen, sollen bereits bekannte Korrelationsmuster repliziert werden: a) positive Korrelation der ERN-Amplitude mit der Fehleranzahl[15,16] und b) negative Korrelation der ERN-Amplitude mit negativem Affekt[19], welcher durch den *Positive and Negative Affect Schedule* (PANAS) Fragebogen erfasst wird. Außerdem soll untersucht werden,

inwieweit die Aufgabe ohne einen signifikanten Verlust an Reliabilität gekürzt werden kann, um die klinische Praktikabilität der Aufgabe zu optimieren.

Die Studie war wie folgt aufgebaut: Für eine Pilotstudie wurden *N*=12 gesunde Probanden rekrutiert, um die Praktikabilität der Stimulusgröße und des adaptiven Reaktionszeitfensters zu testen. In der Hauptanalyse wurden Daten von *N*=14 gesunden Probanden ausgewertet. Alle Teilnehmer wurden mit Hilfe des *Mini International Neuropsychiatric Interview* (M.I.N.I., Deutsche Version 5.0.0[39]) auf psychiatrische Erkrankungen gescreent. Ausschlusskriterien waren aktuelle oder frühere psychiatrische Diagnosen.
Bevor die Testperson die Task bearbeitete, wurde ihr aktueller Affekt mit der deutschen Version des PANAS-Fragebogen erfasst.
Die Probanden bearbeiteten eine modifizierte Version der *Eriksen Flanker Task* (Eriksen & Eriksen, 1979) zweimal im Abstand von 28 Tagen. Bei jedem der Stimuli wurden fünf horizontal ausgerichtete, inkongruente Pfeile auf einem Monitor gezeigt ("<<<><<<" oder ">>><>>>" oder "><><>>" oder "<><><") und die Testperson angewiesen, so schnell und korrekt wie möglich die Richtung des mittleren Pfeils anzugeben. Unmittelbar nach dem Tastendruck bekam der Proband ein Feedback, ob er richtig, falsch oder zu langsam geantwortet hat. Um viele Fehler zu forcieren und damit die Reliabilität der Messung zu maximieren, wurde ein adaptives Reaktionszeitfenster in die Task integriert, welches sich an das Leistungsniveau des Probanden kontinuierlich anpasste.
Die Hirnaktivität wurde mittels EEG mit 64 Elektroden nach dem internationalen 10/20-System aufgezeichnet.

Die Studie zeigt, dass mit der modifizierten *Eriksen Flanker Task* die ERN mit einer exzellenten Reliabilität (ICC=0.806-0.947) gemessen werden kann, was im Vergleich zu vorangegangenen Studien nominal ein besseres Ergebnis ist. Eine mögliche Erklärung dafür ist das adaptive Reaktionszeitfenster, welches zu einer höheren Fehlerrate führt. Mit einer steigenden Anzahl von Fehlern steigt sowohl die ERN-Reliabilität[36] als auch die statistische Power einen signifikanten Unterschied zwischen falschen und richtigen Antworten zu messen[15,40].
Zudem reduziert das adaptive Reaktionszeitfenster Störfaktoren, welche die Reliabilität der ERN negativ beeinflussen könnten. Dazu gehören Lerneffekte zwischen den beiden Messzeitpunkten, welche zu einer Veränderung der ERN-Amplitude führen könnten[14]. Zudem gleicht es unterschiedliche Leistungsniveaus einzelner Probanden aus und kann dadurch einer potenziellen Ergebnisverfälschung bei Gruppenvergleichen entgegenwirken[25,26]. Allerdings muss diesbezüglich bedacht werden, dass das adaptive Reaktionszeitfenster zwar die Fehlerrate vergleichbar macht, doch die individuelle Reaktionszeit einzelner Probanden nicht beeinflusst wird. Dies könnte zu verzerrten Ergebnissen führen[13,41].

Wir fanden Hinweise für die Validität der gemessenen ERN, indem bekannte Korrelationsmuster teilweise repliziert wurden: (A) ein positiver Trend zwischen ERN-Amplitude und der Fehleranzahl[15,16] (r=0.394; p=0.082) und (B) eine signifikant negative Korrelation der ERN-Amplitude mit negativem Affekt[19] (r=-0.583, p=0.014). Dass unsere Studie nur einen Trend zur Signifikanz zwischen ERN-Amplitude und Fehleranzahl zeigen konnte, ist vermutlich auf die kleine Stichprobe zurückzuführen, da unsere gemessene Effektgröße mit den zuvor berichteten Werten übereinstimmt[15]. Das

topographische Mapping der ERN zeigte in Übereinstimmung mit anderen Studien[5,6] maximale Aktivität in frontozentralen Mittellinienelektroden, was die Validität zusätzlich untermauert.
Eine weitere wichtige Fragestellung unsere Studie war, ob die Aufgabe gekürzt werden kann, um die klinische Praktikabilität der Aufgabe zu verbessern. Unsere Analysen lassen den Schluss zu, dass die Aufgabe auf die Hälfte der Trials gekürzt und die ERN innerhalb von ca. 8 Minuten gemessen werden kann, ohne dass die Reliabilität kleiner als 0.8 wird. Dies ermöglicht eine Zeiteffiziente Messung der ERN im klinischen Alltag.

Die Studie legt einen wichtigen Grundstein in der Weiterentwicklung der ERN zum psychiatrischen Biomarker. Es konnte gezeigt werden, dass die ERN mit einer hohen Reliabilität und Validität in einem praktikablen Zeitrahmen gemessen werden kann.
In einem nächsten Schritt sollte unsere modifizierte *Eriksen Flanker Task* in einem balancierten Studiendesign mit einer klassischen Task verglichen und damit der Einfluss des adaptiven Reaktionszeitfensters auf die ERN genauer beurteilt werden. Zudem stellt die Bewertung und Interpretation der ERN eines einzelnen Probanden aufgrund der hohen Varianz verschiedener präanalytischer und analytischer Faktoren[42], die alle einen potenziellen Einfluss auf die Reliabilität haben, ein Problem dar. Zukünftige Studien sollten weitere Kriterien (Geschlecht, Stress, Alter, Vorerkrankungen, Medikamentenwirkungen, zirkadianer Rhythmus etc.) und deren Auswirkungen auf die ERN, insbesondere innerhalb der *Research Domain Criteria* (RDoC)[43]-Matrix[44,45], untersuchen.

7 Overview of manuscripts and publications accepted for publication

Suchan, F., Kopf, J., Althen, H., Reif, A., & Plichta, M. (n.d.). Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task. Acta Neuropsychiatrica, 1-8. doi:10.1017/neu.2018.36

8 Introduction

Distinguishing error from correctness is an essential requirement for learning progress[12]. In order to understand the function of error-related brain activity, an event-related potential (ERP) has been investigated in several electroencephalography (EEG) studies: the *error-related negativity* (ERN), a negative deflection appearing within 100ms after an erroneous response that peaks in fronto-central midline recording sites[5,10]. To elicit the ERN the Eriksen Flanker Task[46] is broadly used[18,32,33,47] which involves discriminating a central target symbol (e.g. an arrow) from surrounding distracting "flanker" symbols. There is strong evidence that the ERN is generated in the anterior cingulate cortex (ACC)[7–9], an area of the medial prefrontal cortex responsible for the integration of affective and cognitive information[48].

A similar, but smaller negative ERP can arise also after correct responses in the same time window and at the same recording sites as the ERN: the correct-related negativity (CRN)[17,18,49,50]. It has been discussed whether the same process[50] or two different processes[51,52] underlie the ERN and CRN.

The function of the ERN is described in different models with regard to an error detection system[10], reinforcement learning[12] or general conflict-detection process[11]. Recently, by application of a *forward model* it has been discovered[13], that the ERN is likely to reflect an error-prediction. This is in line with the *predicted response-outcome* (PRO) model[14], which interprets the ERN as a surprise signal caused by non-occurrence of a predicted event.

Several factors have been shown to influence the ERN. Particularly important is the performance of the individual subject: The higher the error rate, the lower the ERN

amplitude[15,16]. Additionally the structure of the task and the instruction are relevant: a) Using congruent stimuli (i.e. target and flanker arrows point to the same direction) leads to increased ERN compared to incongruent stimuli[17]; b) task instruction focusing on accuracy over speed leads to increased ERN[5,18]; c) the ERN scales with the availability of sensory information and the task goal[53].

Moreover, negative affect[19,20] and several psychiatric disorders[21–24] are related to the ERN amplitude. Recently, it has been demonstrated that the ERN can a) predict the onset of internalizing disorder[25] such as anxiety disorder during the adolescence[26,27], b) provide evidence for therapy responsiveness[21,22,28–30] and c) help to guide treatment decisions[31].

Particularly the latter case emphasizes the clinical relevance of the ERN and makes it a promising candidate as a biomarker for psychiatric disorders. A basic requirement for a biomarker is the reliable measurement. Only a few studies have investigated ERN reliability by using different *Eriksen Flanker Task* variants and found intraclass correlation coefficients (ICCs) between 0.62 and 0.74[32,33,35,36,54].

With the present study we seek to investigate test-retest reliability of the ERN by using a modified *Eriksen Flanker Task* with an adaptive reaction time (RT) deadline[37,38] in two measurement sessions separated by 28 days. The application of an adaptive RT deadline is intended to maximize reliability due to higher error rate[36] while ERN is significantly different from CRN amplitude and a potential decrease of ERN amplitude[15,16] is negligible. At the behavioral level it is expected that the accuracy data is constant across sessions due to the adaptive RT deadline, whereas RT is predicted to be faster in session 2 because of training effects[33]. To ensure the validity of the modified Eriksen Flanker Task, we attempt to replicate known correlation patterns: (1) positive correlation

of ERN amplitude with number of errors[15,16] and (2) negative correlation of ERN amplitude with negative affect, measured by the PANAS questionnaire[19]. In order to optimize a potential future clinical use of the task, we determine whether the task can be shortened without significant loss in reliability.

Aims of the Study

To quantify test-retest reliability of the *error-related negativity* evoked by a modified *Eriksen Flanker Task* with an adaptive reaction time deadline. We seek to determine whether the task can be shortened without significant loss in reliability.

9 Materials and Methods

9.1 Participants

For the pilot study *N*=12 healthy participants were recruited to adjust task parameters. Two subjects had to be excluded from analyses due to technical problems. Power estimation for the main study was calculated based on the pilot study results. Using G*Power 3.1.9.2 we calculated a required sample size of *N*=11 subjects given a statistical power of 0.80, α=0.05 (one-tailed) and an effect size of 0.83 for a *t*-test with dependent means. To compensate for drop-outs a new sample of *N*=15 subjects was recruited for the main study. One subject had to be excluded from the main study due to technical problems. Finally, test-retest data from *N*=14 subjects (9F/5M; mean age=23.5 years, *SD*=2.07 years, range=20-28 years) were included for main analyses. All participants were tested for mental health by the *Mini International Neuropsychiatric Interview* (M.I.N.I.), German Version 5.0.0[55]. Exclusion criteria included current or preceding psychiatric diagnoses. We documented consumption of cigarettes, caffeine (including coffee, coke, or caffeinated tea) and alcohol before the first testing and requested the subjects to appear in a comparable condition for the second testing.

All participants were compensated for their participation and gave written informed consent after detailed explanation of the experimental procedure. The study was approved by the Ethics Committee of the University of Frankfurt and is in accordance with the latest version of the Declaration of Helsinki.

9.2 Positive and Negative Affect Schedule (PANAS)

The German version of the *Positive and Negative Affect Schedule* (PANAS) is a self-report measuring instrument of affect adapted by Krohne, Egloff, Kohlmann, & Tausch, (1996) from the English language questionnaire PANAS[57]. The questionnaire consists of 20 adjectives describing different emotions (see supplementary material). Ten adjectives each cover the dimensions positive affect and negative affect. Every item can be rated on a Five-Point Likert-Type Scale ranging from 1 "not at all" to 5 "extremely". Subjects responded on the basis of their present mood. The sum scores representing negative and positive affect have adequate internal consistency, test–retest reliability, and convergent and discriminant validity[57].

Subjects completed the PANAS before the *Eriksen Flanker Task* started at both sessions. To ensure validity by replicating known correlation patterns, state affect was correlated with the ERN amplitude.

9.3 Modified Eriksen Flanker task

Subjects performed a modified arrow version of the Eriksen Flanker Task[46] two times (session t1 and session t2) separated by exactly 28 days (Figure 1) in an electrically shielded, sound-attenuated, dimly-illuminated room (subjects of the pilot study finished only session t1). Presentation software Version 18.1 (Neurobehavioral Systems, Inc.) was used. The whole task included 411 trials, 12 exercise trials and 399 experimental trials. In order to force many errors only incongruent stimuli were included. On each trial five horizontally aligned arrows were shown in the middle of the monitor (“<<><<“ or “>><>>” or “><><>” or “<><><“) for 125ms followed by a white screen during the RT deadline of maximal 475ms. Each of the stimulus types was intended to be shown

100 times (due to a technical problem, “<<><<“ was only shown 99 times on both sessions and all subjects). The subject was instructed to respond as fast and accurate as possible with the right or left arrow key using his/her right index finger on a keyboard, congruent to the direction of the central arrow. Immediately after the button press, a feedback was presented: a plus (+) sign for correct answers, minus (-) for erroneous answers and exclamation point (!) was shown when the subject did not answer within the current RT deadline. In order to force quick answers, the RT deadline was adjusted after each trial by a reduction of 25ms in case the subject reacted correctly within the current RT deadline or an extension by 25ms in case the response took longer than the current RT deadline. Between each trial a white screen without fixation cross was shown for randomly 500-1500ms.

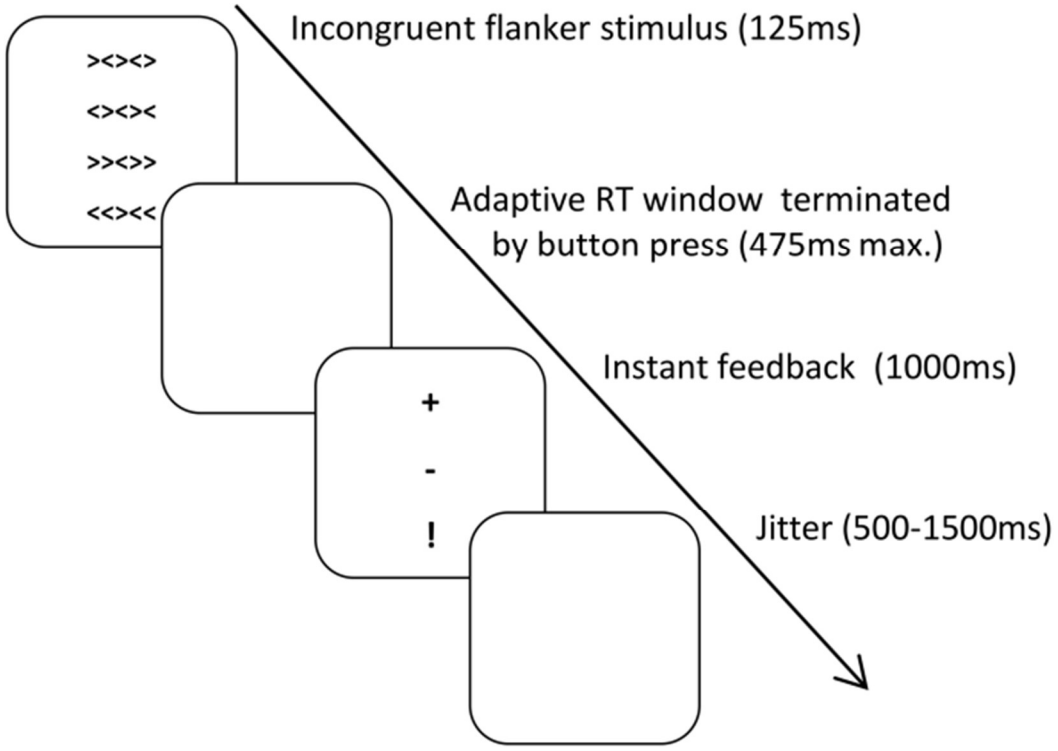


Figure 2 - Execution of the modified Eriksen Flanker Task.

9.4 EEG recording

The EEG was recorded using an elastic head cap with 64 scalp electrodes according to the international 10/20-System. Four additional electrodes were placed to record an electrooculogram (EOG), two close to each angulus oculi lateralis, one on the supercilium and one on the palpebrae inferioris. Ground electrode was placed between the FPz and Fz electrode, reference electrode between the Fz and Cz electrode. All signals were digitized with a 64-channel DC-amplifier and the software “BrainVision Recorder” 2.0 (BrainProducts, Munich, Germany) with a sampling rate of 5000 Hz.

9.5 Data analysis

EEG data were analyzed using the software “BrainVision analyzer” 2.0 (BrainProducts, Munich, Germany). First, electrode TP9 and TP10 were disabled, since they are placed on the mastoid and not used as reference. Data were band pass filtered with a low cutoff of 0.1Hz, a high cutoff of 50Hz and a notch filter of 50Hz. Blinks and eye movements were corrected based on the method established by Gratton, Coles, and Donchin (1983). The algorithm corrects eye artifacts by subtracting the eye channel voltages multiplied by a channel-dependent corrective factor from the respective EEG channels.

Subsequently data were re-referenced on an average reference of all electrodes and the former reference was reused as channel FCz. The EEG was segmented response-locked with an entire length of 800ms, with 400ms pre and post response each. The automatic artifact rejection searched for values exceeding a difference of ±70µV within 200ms and excluded data 200ms before and after the artifact. This procedure did not reveal any artifacts. Afterwards the

segments were averaged separately into correct, error and missed trials and a window -400ms to -200ms prior to the response was used as baseline. The ERP components ERN and CRN were analyzed in terms of area and peak measures at electrode sites FCz and Cz. For area measures the mean activity in the interval 0-100ms after response was calculated, for peak analysis automatic peak detection identified the largest negativity in the same interval.

In the process of our analysis it was necessary to evaluate the EEG data additionally stimulus-locked. The window -400 to -200ms pre-stimulus was used as baseline and the average time course separated into correct and error as well as session t1 and t1 was calculated.

9.6 Statistical methods

For statistical calculations IBM SPSS statistics (version 22) and MATLAB R2017b (The Mathworks, Natick, MA) was used.

In case of behavioral data we used Wilcoxon-test (alpha=0.05; two-tailed) due to non-normally distributed data as tested by Shapiro-Wilk-tests. For differences in peak and area measures of CRN and ERN we used paired *t*-tests (alpha=0.05; two-tailed) after testing for Gaussian distribution by Shapiro-Wilk-tests (all *ps* >0.42).

In order to analyze EEG data, we calculated a 2x2 repeated measure ANOVA with factors (1) accuracy (CRN, ERN) and (2) sessions (t1, t2). Post-hoc dependent *t*-tests (alpha=0.05; two-tailed) were performed in case of significant interaction effects.

Test-retest reliability was assessed by calculating ICC(2,1) for absolute agreement defined by Shrout & Fleiss, (1979) as:

$$ICC(2,1)=BMS-EMS/(BMS+(k-1)*EMS+k*(JMS-EMS)/N)$$

BMS=between-subjects mean square; EMS=error mean square; JMS=session mean square (the original terminology of "J" is "Judge"); *k*=number of repeated sessions and *N*=number of subjects. Thus, in the current study, *k*=2 and *N*=14.

Following Shrout & Fleiss (1979) we defined ICC values <0.4 as poor, 0.4–0.75 as fair to good and >0.75 as excellent. Negative ICC values were reset to 0[60].

For correlation analyses of ERN amplitude, we calculated the correlation according to Spearman (one-tailed), since the scores of negative affect and number of errors were not Gaussian distributed.

10 Results

10.1 Behavioral results

Participants responded significantly faster in session t2 compared to session t1, for both correct and error trials (Table 1). There was a significant effect on number of correct trials, but not on number of error and missed trials between sessions (Table 1). Across sessions the accuracy was consistent at a level of approximately 80% (see supplementary Figure 1).

Table 1 - Performance Data.

	Session t1		Session t2		Session t1 -t2		
	Median	**IQR[1]**	**Median**	**IQR[1]**	**U**	***p***	**dz**
RT correct trials (ms)	404	53	366	36	-3.296	<0.001	1.935
RT error trials (ms)	373	50	343	30	-3.233	<0.001	1.604
Number of correct trials	199	8	201	2	-2.280	0.021	0.565
Number of erroneous trials	68	53	81	55	-0.471	0.659	0.181
Number of missed trials	136	46	120	56	-0.031	0.985	0.017

[1] IQR=interquartil range.

10.2 Comparing CRN and ERN

Figure 2A shows response-locked ERPs for error and correct trials at FCz electrode averaged over all subjects and trials. As expected, there was a significant difference between CRN and ERN (peak amplitude measures: $F(1,13)=16.673$, $p<0.001$, $\eta p^2=0.562$; area measures: $F(1,13)=10.008$, $p=0.007$, $\eta p^2=0.435$) with more pronounced negativity for ERN vs. CRN. For factor session, there was a significant effect (peak amplitude measures: $F(1,13)=15.282$, $p=0.002$, $\eta p^2=0.540$; area measures: $F(1,13)=27.924$, $p<0.001$, $\eta p^2=0.682$) with a

more pronounced negativity for session 1 vs. session 2. Additionally, there was a significant interaction of accuracy and session for peak amplitude measures ($F(1,13)=11.484$, $p=0.005$; $\eta p^2=0.469$) but not for area measures. Post-hoc *t*-tests revealed that the interaction resulted from a significant change in the CRN amplitude ($t=-4.270$, $p<0.001$, $dz=1.141$) while the ERN amplitude difference was not significant across sessions ($t=-1.841$, $p=0.089$, $dz=0.492$).

The topographies showed a more pronounced negativity in frontal areas for error compared to correct trials and the major difference between CRN and ERN in the central cortex (Figure 2B).

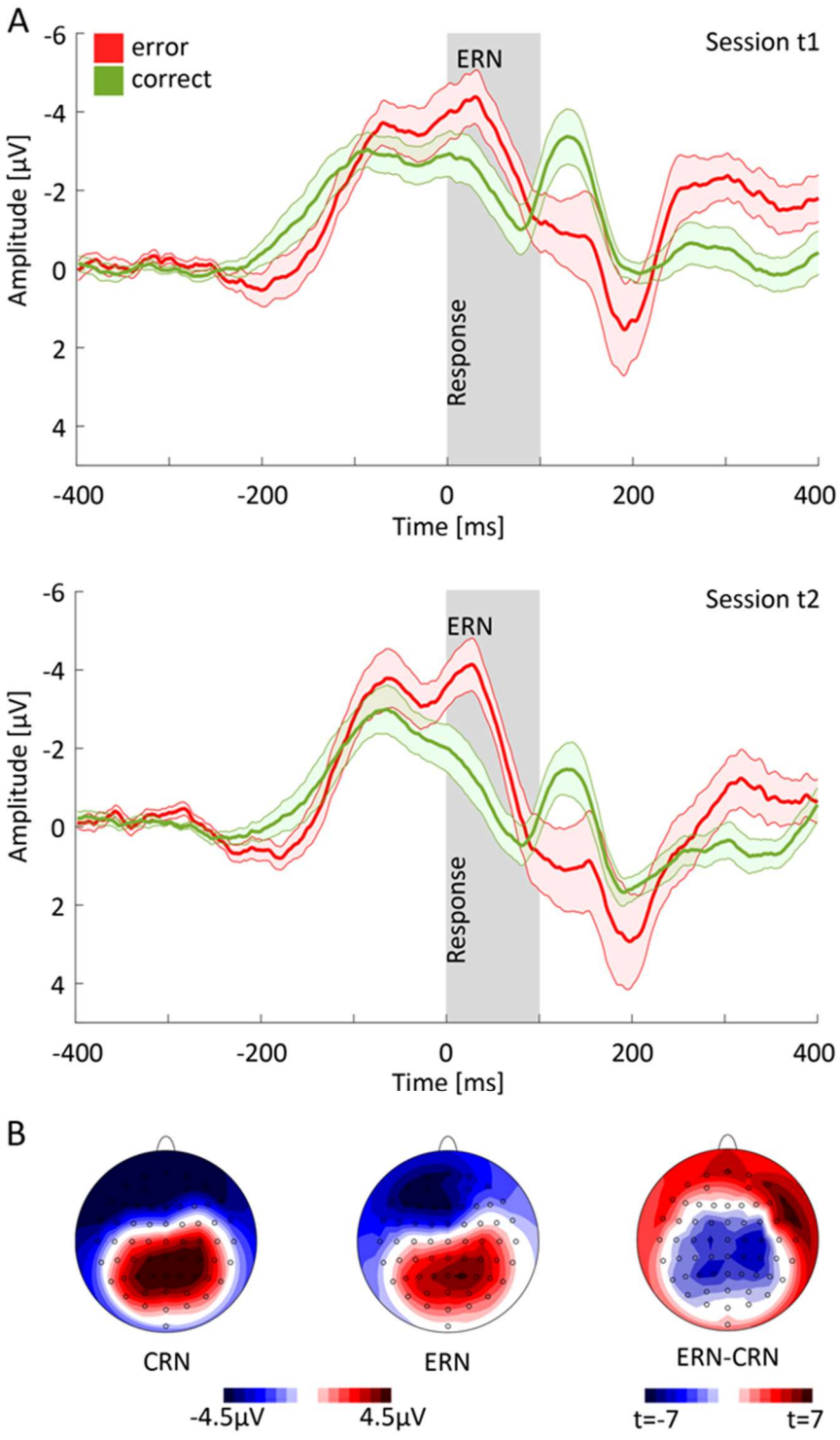


Figure 3 – (a) Response-locked time courses of correct and error trials at FCz electrode (±SE) for session t1 and t2. (b) Topographic mapping of correct-related negativity (CRN) and *error-related negativity* (ERN) and t-map of the difference ERN-CRN (area measure)

10.3 Test-retest reliability

Table 2 shows test-retest reliability indices of ERP measures for error and correct trials at FCz and Cz electrode. Considering the FCz electrode, ICC_{ERN} was excellent (peak amplitude measures: ICC=0.947, p<0.001; area measures: ICC=0.806, p<0.001) and ICC_{CRN} was fair to good (peak amplitude measures: ICC=0.747, p<0.001; area measures: ICC=0.675, p<0.001). For peak amplitude measures the $ICC_{ERN-CRN}$ was excellent (ICC=0.792, p<0.001) and fair to good for area measures (ICC=0.585, p=0.013). On the contrary, peak latency measures were characterized by a low non-significant reliability of ERN (ICC=0.143, p=0.290) and CRN (ICC=0.347, p=0.113) but a moderate and significant reliability of ERN-CRN (ICC=0.690, p=0.002). For Cz electrode we found comparable results.

Table 2 - Test-retest reliability for ERN and CRN at (a) FCz and (b) Cz electrode[1].

(a) FCz electrode:	ERP	ICC[a] (95% CI)	*p*
Peak amplitude measures	ERN	0.947 (0.832-0.983)	<0.001
	CRN	0.747 (0.030-0.930)	<0.001
	ERN-CRN	0.792 (0.226-0.939)	<0.001
Area measures	ERN	0.806 (0.052-0.950)	<0.001
	CRN	0.675 (0.032-0.899)	<0.001
	ERN-CRN	0.585 (0.092-0.846)	0.013
Peak latency measures	CRN	0.347 (0.000-0.736)	0.113
	ERN	0.143 (0.000-0.594)	0.290
	ERN-CRN	0.690 (0.295-0.887)	0.002

(b) Cz electrode:	ERP	ICC[a] (95% CI)	*p*
Peak amplitude measures	ERN	0.890 (0.679-0.964)	<0.001
	CRN	0.740 (0.281-0.914)	<0.001
	ERN-CRN	0.628 (0.195-0.861)	0.004
Area measures	ERN	0.826 (0.482-0.943)	<0.001
	CRN	0.800 (0.482-0.931)	<0.001
	ERN-CRN	0.743 (0.361-0.910)	<0.001
Peak latency measures	ERN	0.569 (0.055-0.840)	0.017
	CRN	0.572 (0.081-0.839)	0.014
	ERN-CRN	0.525 (0.012-0.819)	0.025

[1] Note that the ICCs are comparable at C_2 electrode where the difference between ERN and CRN was at maximum.

[a] ICC for absolute agreement

10.4 Validity

Spearman correlation for the ERN amplitude (FCz) with relative number of errors (Figure 3A) revealed a trend to significance (r=0.394; p=0.082). The correlation of negative affect and ERN amplitude (Figure 3B) reached significance (r=-0.583, p=0.014). Topographic mappings of the correlations show that the absolute maxima were located at central electrodes (Figure 3C).

A negative deflection preceding the ERN is noticeable in our response-locked time courses (Figure 2A). To further examine this negative deflection, we analyzed the EEG data stimulus-locked (see supplementary Figure 2) and identified visual evoked potentials: The negative potential before the ERN and CRN is most likely the N200 which peaks at FCz electrode (correct: 292ms (t1)/281ms (t2); error: 291ms (t1)/ 277ms (t2) post stimulus)[61].

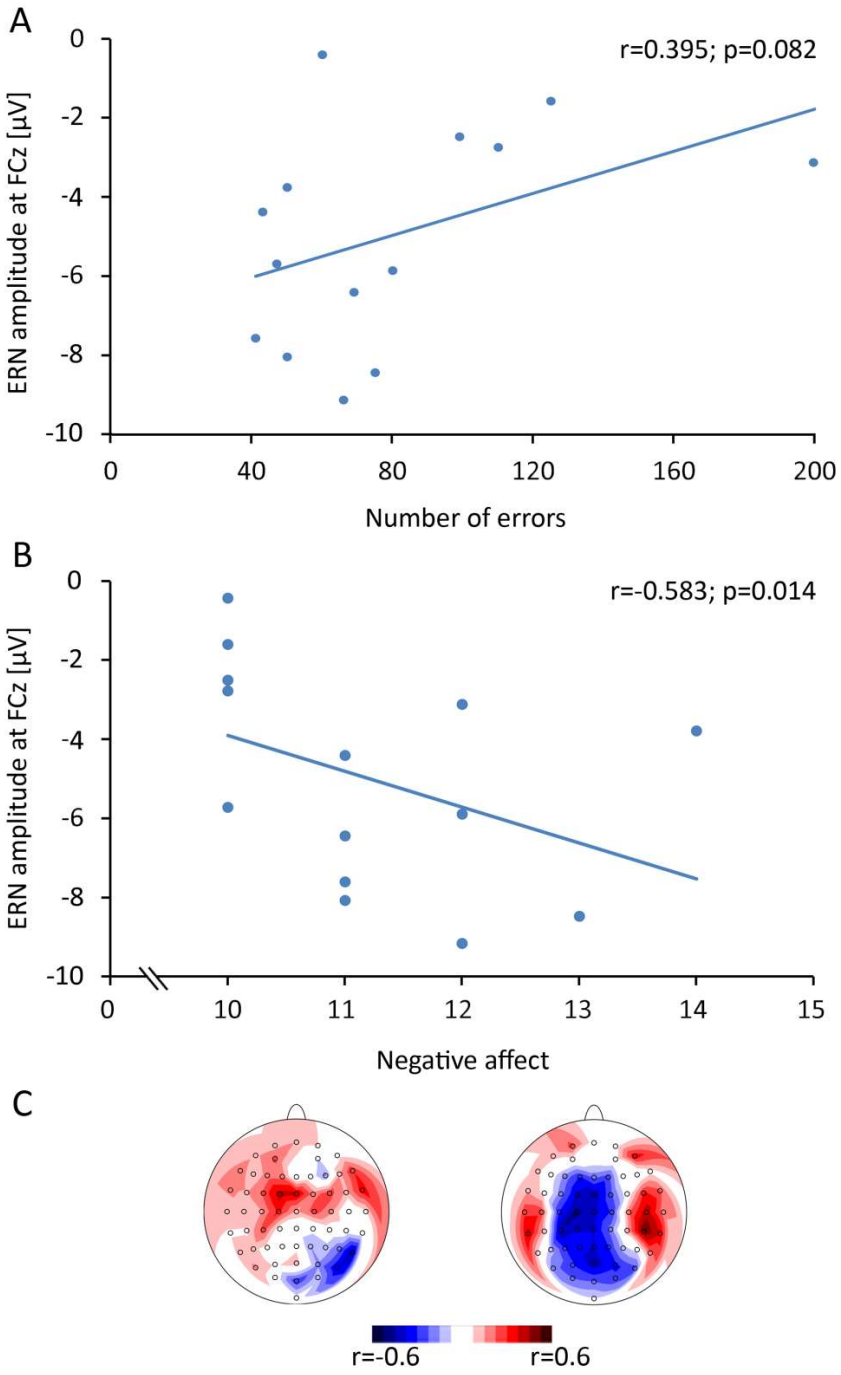


Figure 4 - (a) Correlation of ERN amplitude with absolute number of errors. (b) Correlation of ERN amplitude with negative affect, measured by the PANAS questionnaire. (c) Topographic mapping of correlation values of ERN amplitude with absolute number of errors (left) and negative affect (right).

10.5 Can the task be shortened?

Figure 4 shows ICC_{ERN} and ICC_{CRN} values with increasing number of included trials at FCz electrode. Analyzing peak measures the ICC_{ERN} exceeded the threshold of >0.80 including 35 trials, for area measures 45 trials were required. Analyzing ICC_{CRN} values for peak measures at least 50 trials were required, for area measures the threshold was not exceeded.

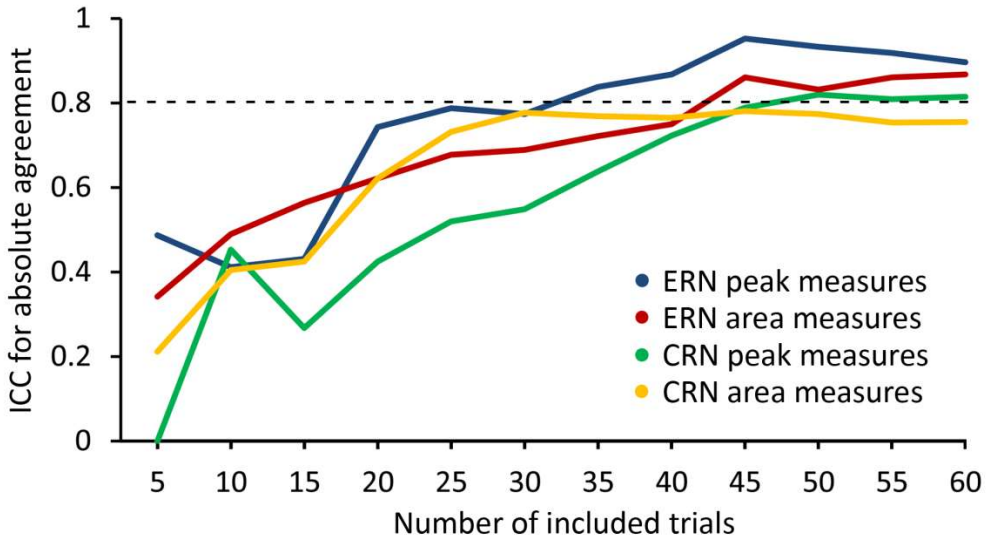


Figure 5 – ICC values of CRN and ERN with increasing number of included trials at FCz electrode.

11 Discussion

The overall objective of this study was to establish ERN acquisition in a reliable, time-efficient and patient-friendly way. Therefore, we used a modified *Eriksen Flanker Task* that increases the number of errors. To ensure external validity we aimed to replicate previously reported correlation patterns of ERN amplitude with number of errors and negative affect. In order to optimize the clinical use of the task, we determined to which extent the task can be shortened while keeping reliability >0.80. Overall, we (A) found excellent reliability of the ERN, which was >0.80 even when the task was reduced to halve of the trials and (B) ensured external validity of the ERN assessed by replicating previously reported correlation patterns with internal and external variables.

11.1 Reliability and effects of the adaptive RT deadline

Excellent reliability of the ERN was found. For peak measures, reliability is higher compared to other studies[32,33,35,36,54], with a 95% CI ranging from 0.832 to 0.983. A potential explanation for this is the adaptive RT deadline which produced about twice as many errors in comparison to other studies. For example, the subjects in the study of Larson et al. (2010) made errors in 12% of the incongruent trials on average. In the studies of Weinberg and Hajcak (2011) and Olvet and Hajcak (2009) error rate was 11.97% and 11.34%. In our paradigm, however, error rate was 20%. (The data refers to the first session, but the second session is comparable). An increasing number of error trials has been shown to increase the ERN reliability[62] and power[15,40].

In addition, the adaptive RT deadline counteracts a potential learning effect. According to the PRO model[14], the ERN amplitude changes with the likelihood of errors. When a

subject performs the same task at two sessions, a learning effect arises and thereby a difference in likelihood for errors between the sessions. However, due to the adaptive RT deadline, the paradigm adapts to the performance level of the subject and the likelihood remains stable despite the learning effect. This may explain the excellent reliability of the ERN as found in our study.

A further advantage of the adaptive RT deadline is performance adjustment across groups. Several studies[15,16] demonstrated a negative relationship between number of errors and ERN amplitude. This can lead to biased results when comparing grou*ps* with different error rates[15]. According to the PRO model[14], different performance levels e.g. in healthy controls and patients would lead to different subjects' expectations of making errors and thus may confound the ERN amplitudes. The adaptive RT deadline can reduce this potential bias because subjects would produce a comparable error rate.

However, there are also potential caveats: a high task performance can be defined not only by the error rate but also by the RT. According to the *forward model*[13,41] better task performance corresponds to more accurate *forward model* predictions about the performance outcome. This could lead to higher ERN amplitudes in subjects with faster RT. Therefore, differences in RT e.g. between patients and healthy controls might lead to biased ERN comparisons.

Finally, other studies generate sufficient number of errors by increasing task length (e.g. 900 trials[36]), while we achieved the high number of errors by a higher error rate. According to the PRO model[14] and Fischer et al., (2017) a smaller ERN amplitude is then expected. However, this ERN amplitude decrement seems to be negligible in our case since we have detected significant differences between ERN and CRN.

11.2 Validity

We found evidence for validity of the recorded ERN by partially replicating known correlation patterns: (A) a trend-wise positive correlation of ERN amplitude with number of errors[15,16] and (B) negative correlation of ERN amplitude with negative affect[19]. In our study, the correlation between ERN amplitude and number of errors showed only a trend to significance. However, this is likely due to the small sample because our revealed effect size is in line with the reported values[15].

An additional aspect supporting the validity of the ERN is the topographic mapping: the ERN peaks in fronto-central midline recording sites as reported by previous studies[5,10]. However, compared to other ERN studies a negative deflection preceding the ERN is noticeable in our response-locked time courses (Figure 2A). To further examine this potential, we evaluated stimulus-locked time courses (see supplementary Figure 2) and identified this negative deflection most likely as the N200. It has been shown in former studies that the N200 appears particularly on incongruent flanker stimuli[61].

11.3 Can the task be shortened?

To determine whether the task can be shortened without significant loss in reliability we analyzed from which number of processed trials a reliability greater than 0.80 can be achieved[63]. Our analyses showed that at least 35 error trials are necessary to achieve reliability >0.80 for peak amplitude measures of the ERN. For area measures, 45 error trials are required. A subject made 68 (t1) and 81 (t2) errors on average during the entire task. Therefore it can be concluded that a reduction of the paradigm to approximately half of the trials (=200) can equally ensure excellent reliability of ERN peak measures. Processing the whole task took on average 16.33

minutes. Thus, our paradigm can acquire highly reliable ERN within 8 minutes. This is advantageous in clinical practice as patients often have shorter concentration spans[64].

11.4 Limitations and Recommendations for future studies

No comparison and reliability assessment of congruent vs. incongruent trials could be conducted, because only incongruent stimuli were shown.

Moreover, our modifications of the Eriksen Flanker Task, i.e. using only incongruent stimuli and an adaptive RT deadline, might have influenced the ERN. For example, it has been shown that faster RTs are associated with larger ERNs[39] while higher error numbers[15,16] and incongruent stimuli[17] lead to reduced ERN amplitudes. In order to investigate these influences systematically, future studies should compare the ERN elicited by a flanker task variant with versus without these modifications.

The instant feedback does not allow for analyzing feedback-related potentials[65]. To achieve this, introducing a delay period between response and feedback would be necessary. Furthermore, contaminations of the response-locked ERP components by the visual feedback cannot be ruled out.

Although sufficient for detecting the ERN with a power >.80, the current sample size does not allow any sub-analyses e.g. gender effects. Studies focusing on such effects should include larger sample sizes.

In order to use the ERN as a biomarker e.g. to control the course of an intervention[31] it is important to assess and interpret the ERN of a single subject (e.g. assignment into treatment type). However, measuring the ERN in single subjects usually is fairly difficult because of high variance due to diverse pre-analytical and analytical sources[42] that all have

an potential impact on reliability. Future studies have to investigate further criteria for establishing the ERN (effects of sex, stress, age, pre-existing disease, medication effects, circadian rhythm etc.) as a trans-diagnostic biomarker in particular within the *Research Domain Criteria* (RDoC)[43] matrix[44,45].

Finally, the task and its practicability should be evaluated in patients to examine feasibility and compare reliability.

11.5 Conclusions

The present study found an excellent reliability of the ERN acquired by a modified *Eriksen Flanker Task* with adaptive RT deadline with only 200 trials which is time-efficient and clinically feasible. Summarizing, the present modified task provides a reliable and efficient recording of the ERN, which will facilitate its use in psychiatry.

12 Darstellung des eigenen Anteils

Die Studie wurde in der Klinik für Psychiatrie, Psychosomatik und Psychotherapie Universitätsklinikum Frankfurt unter Betreuung von Prof. Plichta und Prof. Reif durchgeführt.

Die Konzeption der Studie erfolgte in Zusammenarbeit mit Prof. Plichta und Dr. Kopf.

Sämtliche Daten wurden nach Einarbeitung durch Prof. Plichta und Dr. Kopf von mir eigenständig erhoben.

Die statistische Auswertung, Interpretierung der Ergebnisse und graphische Darstellung erfolgte durch mich mit Unterstützung von Prof. Plichta.

Ich versichere, das Manuskript selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen verwendet zu haben.

13 References

1. Möller, Hans-Jürgen ; Laux, Gerd ; Kapfhammer H-P. *Psychiatrie, Psychosomatik, Psychotherapie : Band 1: Allgemeine Psychiatrie*. Springer Berlin Heidelberg; 2017.
2. Pogarell O. EEG in der Psychiatrie. *Neurophysiologie-Labor*. 2017;39(3):116-128. doi:10.1016/j.neulab.2017.06.005
3. Gallinat J, Mulert C, Leicht G. Stellenwert des Elektroenzephalogramms in der Psychiatrie. *Nervenarzt*. 2016;87(3):323-339. doi:10.1007/s00115-016-0068-2
4. Sur S, Sinha VK. Event-related potential: An overview. *Ind Psychiatry J*. 2009;18(1):70-73. doi:10.4103/0972-6748.57865
5. Gehring, Goss, Coles, Meyer, Donchin. A neural system for error detection and conpensation. *Psychol Sci*. 1993;4(6):385-390.
6. Hohnsbein J, Falkenstein M, Hoormann J, Blanke L. *Effects of Crossmodal Divided Attention on Late ERP Components. I. Simple and Choice Reaction Tasks*. Vol 78.; 1991. doi:10.1016/0013-4694(91)90061-8
7. Brázdil M, Roman R, Daniel P, Rektor I. Intracerebral error-related negativity in a simple Go/NoGo task. *J Psychophysiol*. 2005;19(4):244-255. doi:10.1027/0269-8803.19.4.244
8. Holroyd CB, Dien J, Coles MGH. Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neurosci Lett*. 1998;242(2):65-68. doi:10.1016/S0304-3940(98)00035-4
9. Luu P, Tucker DM, Makeig S. Frontal midline theta and the error-related negativity: Neurophysiological mechanisms of action regulation. *Clin Neurophysiol*. 2004;115(8):1821-1835. doi:10.1016/j.clinph.2004.03.031

10. Falkenstein, Hohnsbein, Hoormann, Blanke. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalogr Clin Neurophysiol*. 1991;78(6):447-455. doi:10.1016/0013-4694(91)90062-9
11. Yeung N, Botvinick MM, Cohen JD. The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol Rev*. 2004;111(4):931-959. doi:10.1037/0033-295X.111.4.931
12. Holroyd CB, Coles MGH. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*. 2002;109(4):679-709. doi:10.1037//0033-295X.109.4.679
13. Joch M, Hegele M, Maurer H, Müller H, Maurer LK. Brain negativity as an indicator of predictive error processing: the contribution of visual action effect monitoring. *J Neurophysiol*. 2017;118(1):486-495. doi:10.1152/jn.00036.2017
14. Alexander WH, Brown JW. Medial prefrontal cortex as an action-outcome predictor. *Nat Neurosci*. 2011;14(10):1338-1344. doi:10.1038/nn.2921
15. Fischer AG, Klein TA, Ullsperger M. Comparing the error-related negativity across groups: The impact of error- and trial-number differences. *Psychophysiology*. 2017;54(7):998-1009. doi:10.1111/psyp.12863
16. Hajcak G, McDonald N, Simons RF. To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*. 2003;40(6):895-903. doi:10.1111/1469-8986.00107
17. Scheffers MK, Coles MGH. Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *J Exp Psychol Hum Percept Perform*. 2000;26(1):141-151. doi:10.1037/0096-1523.26.1.141

18. Falkenstein, Hoormann, Christ, Hohnsbein. ERP components on reaction errors and their functional significance: A tutorial. *Biol Psychol*. 2000;51(2-3):87-107. doi:10.1016/S0301-0511(99)00031-9
19. Hajcak G, McDonald N, Simons RF. Error-related psychophysiology and negative affect. *Brain Cogn*. 2004;56(2 SPEC. ISS.):189-197. doi:10.1016/j.bandc.2003.11.001
20. Luu P, Collins P, Tucker DM. Mood, personality, and self-monitoring: Negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J Exp Psychol Gen*. 2000;129(1):43-60. doi:10.1037/0096-3445.129.1.43
21. Fissler M, Winnebeck E, Schroeter TA, et al. Brief training in mindfulness may normalize a blunted error-related negativity in chronically depressed patients. *Cogn Affect Behav Neurosci*. 2017;17(6):1164-1175. doi:10.3758/s13415-017-0540-x
22. Rabella M, Grasa E, Corripio I, et al. Neurophysiological evidence of impaired self-monitoring in schizotypal personality disorder and its reversal by dopaminergic antagonism. *NeuroImage Clin*. 2016;11:770-779. doi:10.1016/j.nicl.2016.05.019
23. Gehring WJ, Himle J, Nisenson LG. Action-Monitoring Dysfunction in Obsessive-Compulsive Disorder Author (s): William J . Gehring , Joseph Himle and Laura G . Nisenson Published by : Sage Publications , Inc . on behalf of the Association for Psychological Science Stable URL : http://www. 2016;11(1):1-6.
24. Meyer A, Hajcak G, Glenn CR, Kujawa AJ, Klein DN. Error-related brain activity is related to aversive potentiation of the startle response in children, but only the ern is associated with anxiety disorders. *Emotion*. 2017;17(3):487-496. doi:10.1037/emo0000243

25. Meyer A, Danielson CK, Danzig AP, et al. Neural Biomarker and Early Temperament Predict Increased Internalizing Symptoms After a Natural Disaster. *J Am Acad Child Adolesc Psychiatry*. 2017;56(5):410-416. doi:10.1016/j.jaac.2017.02.005
26. Meyer A. A biomarker of anxiety in children and adolescents: A review focusing on the error-related negativity (ERN) and anxiety across development. *Dev Cogn Neurosci*. 2017;27(March):58-68. doi:10.1016/j.dcn.2017.08.001
27. Meyer A, Nelson B, Perlman G, Klein DN, Kotov R. A neural biomarker , the error-related negativity , predicts the first onset of generalized anxiety disorder in a large sample of adolescent females. 2018. doi:10.1111/jcpp.12922
28. Schroder HS, Moran TP, Moser JS. The effect of expressive writing on the error-related negativity among individuals with chronic worry. *Psychophysiology*. 2018;55(2). doi:10.1111/psyp.12990
29. Forster SE, Zirnheld P, Shekhar A, Steinhauer SR, O'Donnell BF, Hetrick WP. Event-related potentials reflect impaired temporal interval learning following haloperidol administration. *Psychopharmacology (Berl)*. 2017;234(17):2545-2562. doi:10.1007/s00213-017-4645-2
30. Hobson NM, Bonk D, Inzlicht M. Rituals decrease the neural response to performance failure. *PeerJ*. 2017;5:e3363. doi:10.7717/peerj.3363
31. Gorka SM, Burkhouse KL, Klumpp H, et al. Error-Related Brain Activity as a Treatment Moderator and Index of Symptom Change during Cognitive-Behavioral Therapy or Selective Serotonin Reuptake Inhibitors. *Neuropsychopharmacology*. 2017;43(6):1355-1363. doi:10.1038/npp.2017.289
32. Cassidy SM, Robertson IH, O'Connell RG. Retest

reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*. 2012;49(5):659-664. doi:10.1111/j.1469-8986.2011.01349.x

33. Olvet DM, Hajcak G. Reliability of error-related brain activity. *Brain Res*. 2009;1284:89-99. doi:10.1016/j.brainres.2009.05.079
34. Segalowitz SJ, Santesso DL, Murphy TI, Homan D, Chantziantoniou DK, Khan S. Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*. 2010;47(2):260-270. doi:10.1111/j.1469-8986.2009.00942.x
35. Weinberg A, Hajcak G. Longer term test-retest reliability of error-related brain activity. *Psychophysiology*. 2011;48(10):1420-1425. doi:10.1111/j.1469-8986.2011.01206.x
36. Larson MJ, Baldwin SA, Good DA, Fair JE. Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. *Psychophysiology*. 2010;47(6):no-no. doi:10.1111/j.1469-8986.2010.01022.x
37. Debener S. Trial-by-Trial Coupling of Concurrent Electroencephalogram and Functional Magnetic Resonance Imaging Identifies the Dynamics of Performance Monitoring. *J Neurosci*. 2005;25(50):11730-11737. doi:10.1523/JNEUROSCI.3286-05.2005
38. Unger K, Heintz S, Kray J. Punishment sensitivity modulates the processing of negative feedback but not error-induced learning. *Front Hum Neurosci*. 2012;6(June):1-16. doi:10.3389/fnhum.2012.00186
39. Quik EH. *The Somatotropic Axis : Effects on Brain and Cognitive Functions*.; 2012.
40. Boudewyn MA, Luck SJ, Farrens JL, Kappenman ES. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*. 2017;(September):e13049.

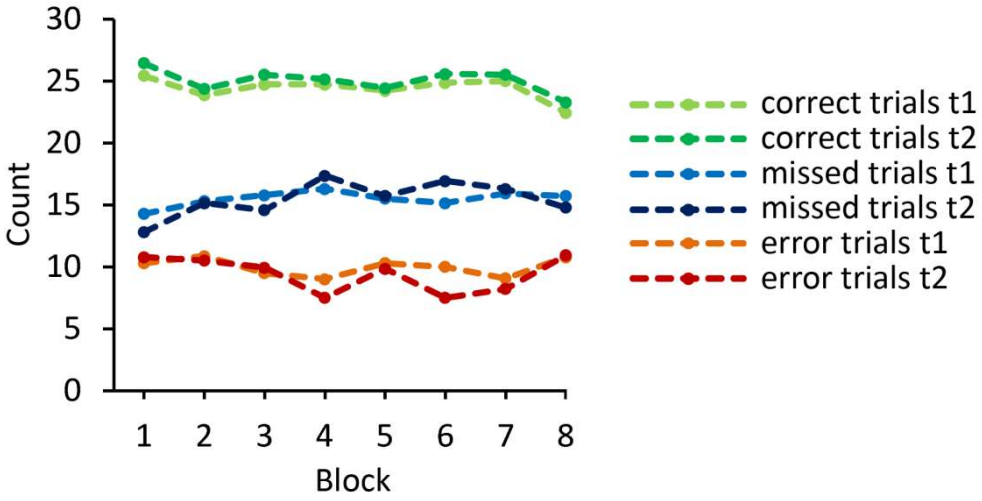
doi:10.1111/psyp.13049

41. Joch M, Hegele M, Maurer H, Müller H, Maurer LK. Accuracy of Motor Error Predictions for Different Sensory Signals. *Front Psychol*. 2018;9(August):1-13. doi:10.3389/fpsyg.2018.01376
42. Micheel CM, Ball JR. *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*.; 2010. doi:10.17226/12869
43. Carcone D, Ruocco AC. Six Years of Research on the National Institute of Mental Health's Research Domain Criteria (RDoC) Initiative: A Systematic Review. *Front Cell Neurosci*. 2017;11(March):1-8. doi:10.3389/fncel.2017.00046
44. Ladouceur CD. The error-related negativity: A transdiagnostic marker of sustained threat? *Psychophysiology*. 2016;53(3):389-392. doi:10.1111/psyp.12585
45. Weinberg A, Meyer A, Hale-Rude E, et al. Error-related negativity (ERN) and sustained threat: Conceptual framework and empirical evaluation in an adolescent sample. *Psychophysiology*. 2016;53(3):372-385. doi:10.1111/psyp.12538
46. Eriksen CW, Eriksen BA. Target redondancy in visual search: Do repetitions of target within the display impair processing? *Percept Psychophys*. 1979;26(3):356-370. https://link.springer.com/content/pdf/10.3758%2FBF03199869.pdf.
47. Ehlis AC, Herrmann MJ, Bernhard A, Fallgatter AJ. Monitoring of internal and external error signals. *J Psychophysiol*. 2005;19(4):263-269. doi:10.1027/0269-8803.19.4.263
48. Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior cingulate cortex. *TrendsCogn Sci*. 2000;4(6):215-222. doi:10.1016/S1364-6613(00)01483-2

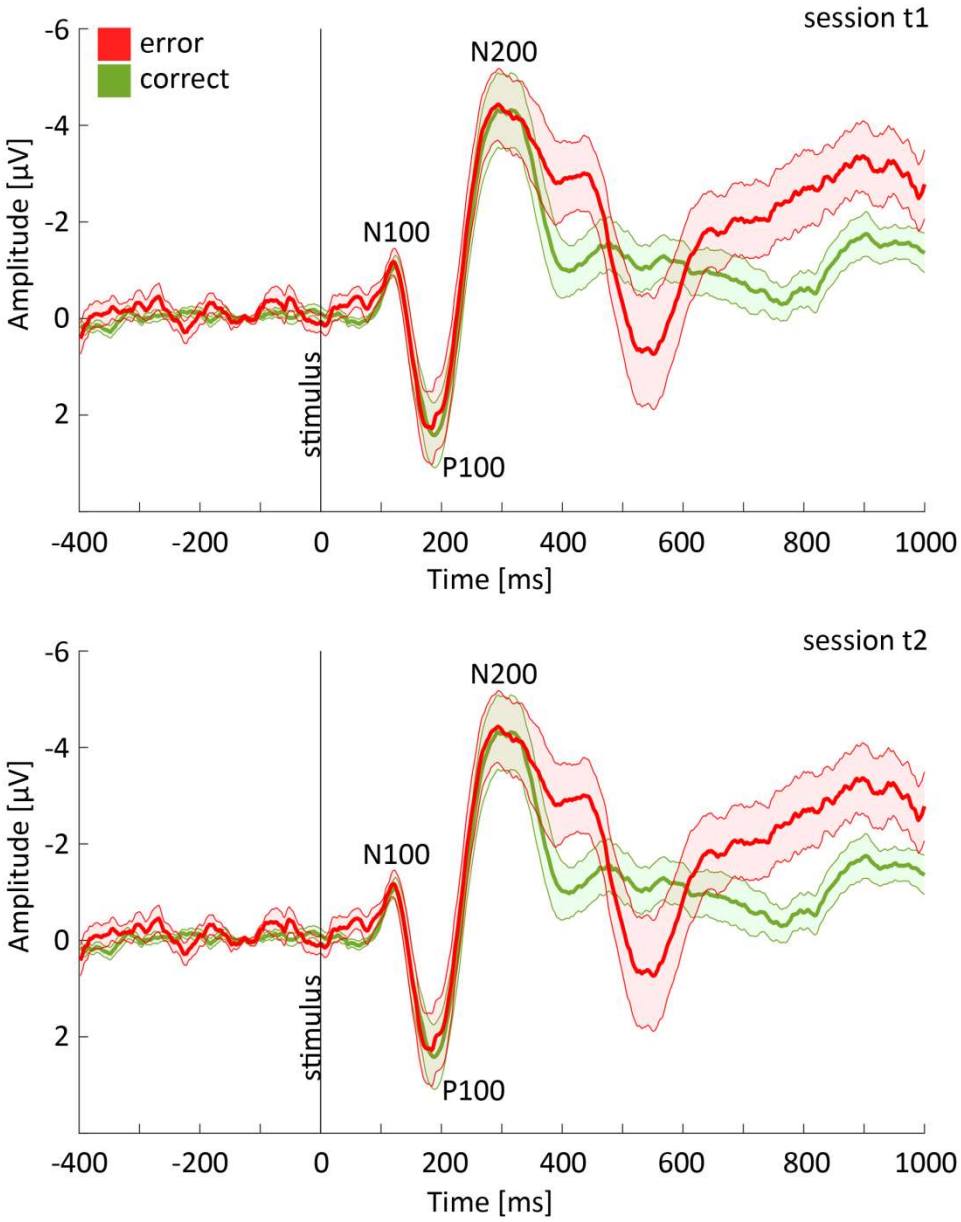
49. Gehring WJ, Knight RT. Prefrontal-cingulate interactions in action monitoring. *Nat Neurosci*. 2000;3(5):516-520. doi:10.1038/74899
50. Vidal F, Hasbroucq T, Grapperon J, Bonnet M. Is the "error negativity" specific to errors? *Biol Psychol*. 2000;51(2-3):109-128. doi:10.1016/S0301-0511(99)00032-0
51. Coles MGH, Scheffers MK, Holroyd CB. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biol Psychol*. 2001;56(3):173-189. doi:10.1016/S0301-0511(01)00076-X
52. Yordanova J, Falkenstein M, Hohnsbein J, Kolev V. Parallel systems of error processing in the brain. *Neuroimage*. 2004;22(2):590-602. doi:10.1016/j.neuroimage.2004.01.040
53. Brown JW, Braver TS. Learned predictions of error likelihood in the anterior cingulate cortex. *Science (80-)*. 2005;307(5712):1118-1121. doi:10.1126/science.1105783
54. Segalowitz SJ, Santesso DL, Murphy TI, Homan D, Chantziantoniou DK, Khan S. Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*. 2010;47(2):260-270. doi:10.1111/j.1469-8986.2009.00942.x
55. Sheehan D V., Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998;59(SUPPL. 20):22-33. doi:10.1016/S0924-9338(99)80239-9
56. Krohne H, Egloff B, Kohlmann C-W, Tausch A. *Untersuchungen Mit Einer Deutschen Version Der "Positive and Negative Affect Schedule" (PANAS)*. Vol 42.; 1996. doi:10.1037/t49650-000

57. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J Pers Soc Psychol*. 1988;54(6):1063-1070. doi:10.1037/0022-3514.54.6.1063
58. Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol*. 1983;55(4):468-484. doi:10.1016/0013-4694(83)90135-9
59. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. http://www.ncbi.nlm.nih.gov/pubmed/18839484.
60. Bartko JJ. On various intraclass correlation reliability coefficients. *Psychol Bull*. 1976;83(5):762-765. doi:10.1037/0033-2909.83.5.762
61. Kopp B, Rist F, Mattler U. N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*. 1996;33(3):282-294. doi:10.1111/j.1469-8986.1996.tb00425.x
62. Larson MJ, Baldwin SA, Good DA, Fair JE. Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. *Psychophysiology*. 2010;47(6):1167-1171. doi:10.1111/j.1469-8986.2010.01022.x
63. Rosaroso RC. Using Reliability Measures in Test Validation. *Eur Sci J*. 2015;11(18):1857-7881. https://eujournal.org/index.php/esj/article/viewFile/5847/5662.
64. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*.; 2013. doi:10.1176/appi.books.9780890425596.744053
65. Bismark AW, Hajcak G, Whitworth NM, Allen JJB. The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology*. 2013;50(2):125-133. doi:10.1111/j.1469-

14 Supplements



Supplemetary Figure 1 - Distribution of correct, missed and error trials over the experimental time course. Each block contains 50 trials, except block 8 contains 49 trials.



Supplemetary Figure 2 - Stimulus-locked time courses of correct and error trials at FCz electrode (+/-SE) for session t1 and t2.

PANAS

Dieser Fragebogen enthält eine Reihe von Wörtern, die unterschiedliche Gefühle und Empfindungen beschreiben. Lesen Sie jedes Wort und tragen Sie dann in die Skala neben jedem Wort die Intensität ein. Sie haben die Möglichkeit, zwischen fünf Abstufungen zu wählen:

1. ganz wenig oder gar nicht 2. ein bisschen 3. einigermaßen 4. erheblich 5.

äußerst Geben Sie bitte an, wie Sie sich <u>im Moment</u> fühlen.

	ganz wenig oder gar nicht	ein bisschen	einiger-maßen	erheblich	äußerst
aktiv	O	O	O	O	O
bekümmert	O	O	O	O	O
interessiert	O	O	O	O	O
freudig erregt	O	O	O	O	O
verärgert	O	O	O	O	O
stark	O	O	O	O	O
schuldig	O	O	O	O	O
erschrocken	O	O	O	O	O
feindselig	O	O	O	O	O
angeregt	O	O	O	O	O
stolz	O	O	O	O	O
gereizt	O	O	O	O	O
begeistert	O	O	O	O	O
beschämt	O	O	O	O	O
wach	O	O	O	O	O
nervös	O	O	O	O	O
entschlossen	O	O	O	O	O
aufmerksam	O	O	O	O	O
durcheinander	O	O	O	O	O
ängstlich	O	O	O	O	O

Schriftliche Erklärung

Ich erkläre ehrenwörtlich, dass ich die dem Fachbereich Medizin der Johann Wolfgang Goethe-Universität Frankfurt am Main zur Promotionsprüfung eingereichte Dissertation mit dem Titel

Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task.

in der Klinik für Psychiatrie, Psychosomatik und Psychotherapie, Universitätsklinikum Frankfurt unter Betreuung und Anleitung von Prof. Plichta ohne sonstige Hilfe selbst durchgeführt und bei der Abfassung der Arbeit keine anderen als die in der Dissertation angeführten Hilfsmittel benutzt habe. Darüber hinaus versichere ich, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Ich habe bisher an keiner in- oder ausländischen Universität ein Gesuch um Zulassung zur Promotion eingereicht. Die vorliegende Arbeit wurde bisher nicht als Dissertation eingereicht.

Vorliegende Ergebnisse der Arbeit wurden in folgendem Publikationsorgan veröffentlicht:
Suchan, F., Kopf, J., Althen, H., Reif, A., & Plichta, M. (n.d.). Reliable and efficient recording of the error-related negativity with a speeded Eriksen Flanker Task. Acta Neuropsychiatrica, 1-8. doi:10.1017/neu.2018.36

(Ort, Datum) (Unterschrift)