International Journal of Psychophysiology 215 (2025) 113236

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Review

Decoding deception with the P300: A meta-analysis of the Concealed Information Test

Julia Knappe[a], Markus Ullsperger[a,b,c], Hans Kirschner[a,*]

[a] Institute of Psychology, Otto-von-Guericke University, D-39106 Magdeburg, Germany
[b] Center for Behavioral Brain Sciences, D-39106 Magdeburg, Germany
[c] German Center for Mental Health (DZPG), partner site Halle-Jena-Magdeburg, D-39106 Magdeburg, Germany

ARTICLE INFO

Keywords:
Concealed Information Test (CIT)
Guilty Knowledge Test (GKT)
Event-related potential (ERP)
P300
Meta-analysis
Forensic psychology

ABSTRACT

The Concealed Information Test (CIT) is frequently used to determine the presence of crime-related information in a suspect's memory. In this paper, we conducted a meta-analysis to test the validity of the CIT to differentiate between guilty and innocent individuals based on amplitude differences of the P300 component of the event-related potential. We included $k = 54$ experimental studies that used either the mock-crime paradigm or the personal-item paradigm. The results show a large mean effect size (d^*) of *1.59* for the P300. Moderation analysis showed that P300 effects in CIT are affected by the choice of paradigm (personal-item vs. mock-crime paradigm), the chosen trial protocol (complex vs. original) and the likelihood of subjects to employ countermeasures. Based on our findings, we conclude that the P300 is useful to determine the presence of crime-related information and that people interested in using the CIT should use the complex trial protocol to maximize effect sizes.

1. Introduction

The detection of deception has a long history, but even trained experts rarely perform above chance level when judging honesty (Bond Jr. and DePaulo, 2006). This limitation has led to growing interest in developing objective methods to detect concealed knowledge using psychophysiological and neuroscientific approaches. One such method is the Concealed Information Test (CIT), also referred to as the Guilty Knowledge Test (GKT), which is grounded in solid cognitive psychophysiological theory (Iacono, 2007; Verschuere and Ben-Shakhar, 2011). The CIT is designed to assess whether a suspect recognizes specific crime-related information. In a typical CIT, multiple-choice items contain one critical probe item and several neutral, irrelevant alternatives. The assumption is that only guilty individuals will consistently show stronger physiological responses to the probe (Lykken, 1959, 1998). Early studies confirmed that recognition of relevant stimuli elicits stronger orienting responses (e.g., Gati and Ben-Shakhar, 1990; Siddle, 1991; Sokolov, 1963). Importantly, the goal of the CIT is not to detect lies in a general sense, but to infer whether the suspect possesses knowledge known only to the perpetrator and law enforcement. This inference is based on involuntary physiological responses, not on verbal reports. Traditionally, the CIT has used autonomic nervous system (ANS) measures such as skin conductance, heart rate, or respiration. More recently, researchers have increasingly focused on event-related brain potentials (ERPs), particularly the P300 component, as a more direct indicator of stimulus recognition (Polich, 2007). The P300 is a centroparietal positive deflection that occurs approximately 300 to 800 milliseconds after stimulus presentation and reflects the perceived relevance of a stimulus. Compared to ANS-based measures, EEG provides direct access to neural-level processing rather than peripheral bodily reactions. It also allows for shorter inter-stimulus intervals, thereby increasing the efficiency of testing. Although the P300 shows habituation effects (Polich, 1989), these are less pronounced than those observed with ANS measures (Rushby and Barry, 2009), which further supports EEG as a promising method in forensic applications.

Meijer et al. (2014) conducted a meta-analysis on the validity of the CIT to determine the presence or absence of crime-related information in a suspect's memory based on four physiological measures. They included studies that either used skin conductance responses (SCR), respiration line length (RLL), heart rate (HR), or amplitudes of the P300 component of the event-related potential (P300) as physiological responses to the stimuli. Results revealed large effect sizes for all measures with the P300 outperforming to autonomic nervous system (ANS) measures. In a related meta-analysis, Leue and Beauducel (2019) examined P300

[*] Corresponding author.
E-mail addresses: markus.ullsperger@ovgu.de (M. Ullsperger), hans.kirschner@ovgu.de (H. Kirschner).

https://doi.org/10.1016/j.ijpsycho.2025.113236
Received 14 May 2025; Received in revised form 3 July 2025; Accepted 30 July 2025
Available online 5 August 2025

amplitudes in relation to deception across both legal and social contexts. Their work focused on the cognitive processes underlying deception tasks and identified two key mechanisms reflected in the P300 component. According to the salience account, P300 differences are thought to reflect increased cognitive engagement with concealed, familiar stimuli compared to unfamiliar ones (Kok, 2001). In contrast, the mental effort account links P300 differences to the cognitive effort required to suppress known information (Beauducel et al., 2006). It is likely that both processes contribute jointly to the P300 response observed in deception tasks (Leue and Beauducel, 2019). In contrast to Meijer et al. (2014), Leue and Beauducel (2019) analysed a broader variety of deception paradigms beyond the CIT, including tasks with lower stakes (e.g. responses have no direct consequences for the subject) or social-evaluative contexts. While this broader perspective provides valuable insights into the generalizability of deception-related P300 responses, it is less focused on the forensic application of the CIT. Moreover, other meta-analyses using fMRI have investigated the neural correlates of deception (Christ et al., 2008; Garrigan et al., 2016). While these studies have provided valuable insights into the brain regions involved in deception, the high monetary cost of fMRI and contraindications for a substantial portion of the population make it unsuitable for day-to-day forensic practice.

The present study aims to provide a comprehensive overview of the current state of research on the P300-based Concealed Information Test (CIT) studies, substantially extending prior evaluations (e.g., Leue and Beauducel, 2019; Meijer et al., 2014). By incorporating 16 additional studies and applying advanced statistical techniques, including mixed-effects meta-regression, we evaluate the effectiveness of the CIT in distinguishing between guilty and innocent individuals based on P300 amplitude differences. In contrast to previous reviews, we systematically examine a broad set of methodological moderators—such as paradigm type (mock-crime vs. personal-item), trial protocol (three-stimulus vs. complex), use of countermeasures, inclusion of control participants, number of CIT questions, and year of publication—and explore their potential interaction effects. Our goal was to clarify the conditions under which the P300-based CIT is most effective and to provide evidence-based recommendations for its optimized use in forensic practice. The key findings are detailed in the results section below and summarized in a schematic figure presented in the discussion section (Fig. 5).

2. Methods

2.1. Study selection procedure

Several methods were employed for literature research. First, we conducted an electronic literature search in the databases scholar.google.com, elicit.org, PubMed and Scopus. We included studies that were available online until the end of December 2024. The literature search was performed using Guilty Knowledge Test, Concealed Information Test and P300 as keywords (for specific search queries see Table 1) and limited to articles written in German or English.

Additionally, all relevant studies meeting the criteria from Meijer et al. (2014) were included in the study set. Finally, we also consulted the reference lists of related reviews (Rosenfeld, 2020; Verschuere and Ben-Shakhar, 2011; Vrij, 2008) as well as the meta-analysis by Leue and Beauducel (2019). Inclusion criteria closely followed Meijer et al. (2014), with a focus on P300-based CIT studies considered relevant for this work. This procedure resulted in a total of 90 unique studies or conditions, of which $k = 54$ were included in the analysis (see Fig. 1). All excluded studies, along with exclusion justifications are listed in the supplementary information. The inclusion criteria were as follows: a study or experimental condition within a study was included if it used either the personal-item or mock-crime paradigm within a peak-to-peak method P300-based CIT and fulfilled the following criteria: (a) It included at least one set of equivalent alternative items, one of which was the relevant (probe) item. (b) The proportion of relevant alternatives (probe items) in a block did not exceed $p = 0.33$ (referred to as BR, as in Meijer et al. (2014)). As previously described, a larger proportion of probe items than $p = 0.33$ is very uncharacteristic in the application of CIT, considering the required relevance and rarity of occurring stimuli for P300 according to Polich (2007). Studies reporting higher proportions were therefore excluded. (c) The CIT was conducted under standard conditions. Studies were excluded if they required participants to imagine, see, or plan a crime instead of committing it. Additionally, studies using a different, modified form of CTP or included practice sessions were excluded. Studies were not considered if they involved non-standardized responses in the CIT task (e.g., instructing participants to respond deceptively to 50 % of stimuli). Studies were also excluded if they provided participants with feedback on their performance or if innocent participants received information about crime-related items. (d) The CIT study provided a measure of accuracy or differentiation between responses triggered by relevant items from guilty and innocent participants. Studies without provided or obtainable information on this aspect, as in Meijer et al. (2014), led to the exclusion of studies. Studies were also excluded if they adopted conditions or control group results from earlier studies. Furthermore, studies were not considered if they exhibited perfect hit rates in line with Meijer et al. (2014).

It is worth noting that both, studies with significant and non-significant results were included (e.g., Mertens and Allen, 2008). The rationale is based on the assumption that P300 is a good indicator of detecting hidden information. However, the amplitude is dependent on both the chosen protocol and potential countermeasures. Hence, non-significant results, i.e., no reliable assignment to the respective group of guilty and innocent individuals, can occur due to these and other variables.

Table 1
Specific queries used for the literature search.

Source	Search query
Google Scholar	"P300" AND "Concealed Information Test" OR "Guilty Knowledge Test" AND "Concealed Information Test" OR "P300"
Elicit.org	"P300-based concealed information detection" OR "ERP lie detection" AND "CIT"
PubMed	((Concealed Information Test) AND (P300)) OR ((Guilty Knowledge Test) AND (P300))
Scopus	((Concealed Information Test) AND (P300)) OR ((Guilty Knowledge Test) AND (P300))

2.2. Moderating factors in P300-based Concealed Information Test (CIT)

2.2.1. Paradigm

There are two P300 implementations of the CIT. First, the mock-crime paradigm ((Lykken, 1959)). Here, participants in the guilty condition commit a mock-crime. Thereafter, participants are presented with both a probe item (crime-related object) and several neutral (irrelevant) control alternatives, and they are instructed to respond "no" for "not known" to all items. Second, the personal-item paradigm. In this paradigm participant's personal items (e.g., first name, last name, birthdate) are used as probe items, embedded among several neutral (irrelevant) control items of the same category (e.g., randomized first name, last name, date). Again, participants are instructed to respond "no" for "not known" to all items.

The results from Meijer et al. (2014) suggest that the personal-item paradigm was associated with a significantly larger P300 amplitude difference between the control group and the "guilty condition" compared to the mock-crime paradigm. The authors suggest as a possible explanation for this that the probe items in studies using the personal-item paradigm could be more salient than in studies using the mock-crime paradigm. Interestingly, they note that the effect size in both paradigms could potentially be enhanced using updated trial protocols (see below, Rosenfeld et al. (2008)).

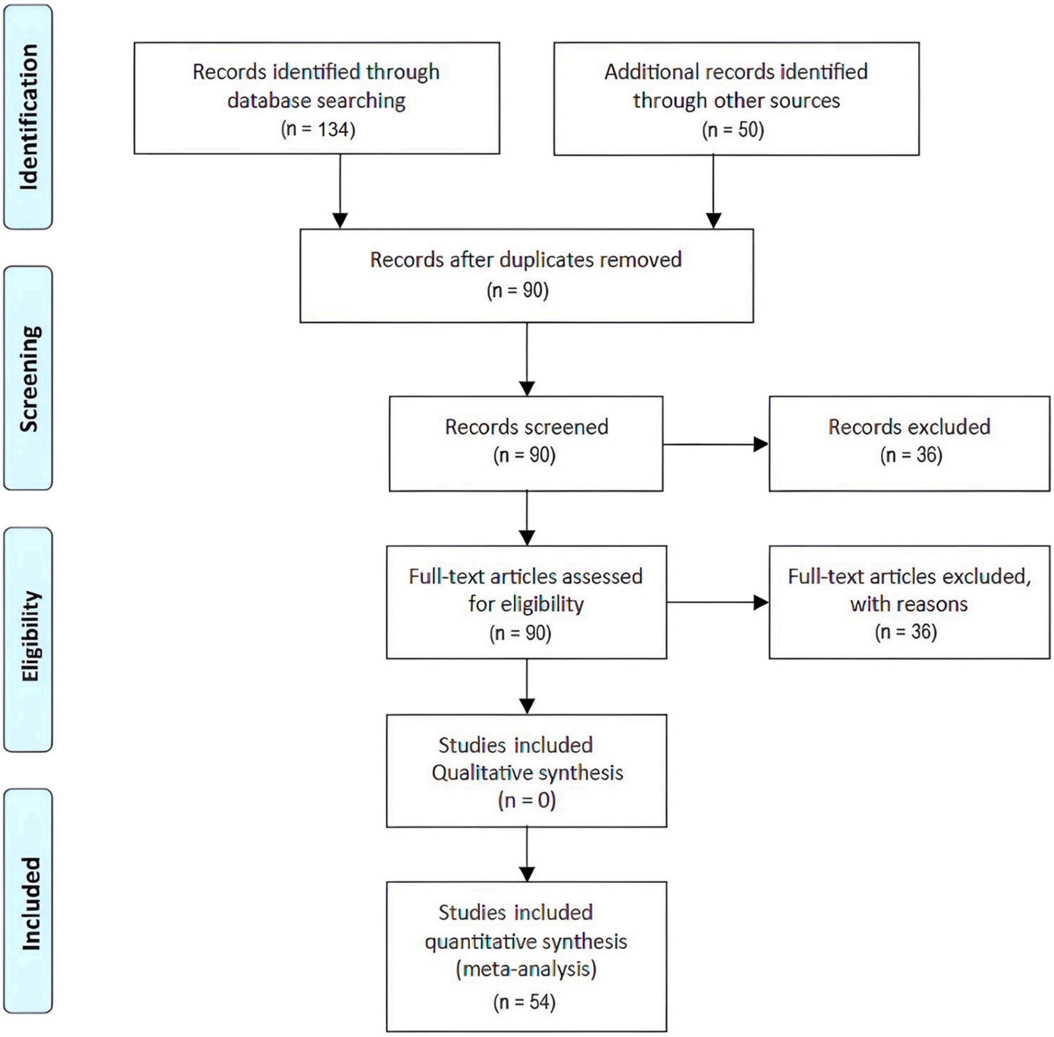

Identification
Screening
Eligibility
Included
Records identified through database searching (n = 134)
Additional records identified through other sources (n = 50)
Records after duplicates removed (n = 90)
Records screened (n = 90)
Records excluded (n = 36)
Full-text articles assessed for eligibility (n = 90)
Full-text articles excluded, with reasons (n = 36)
Studies included Qualitative synthesis (n = 0)
Studies included quantitative synthesis (meta-analysis) (n = 54)

Fig. 1. PRISMA flow diagram according to Moher et al. (2009).

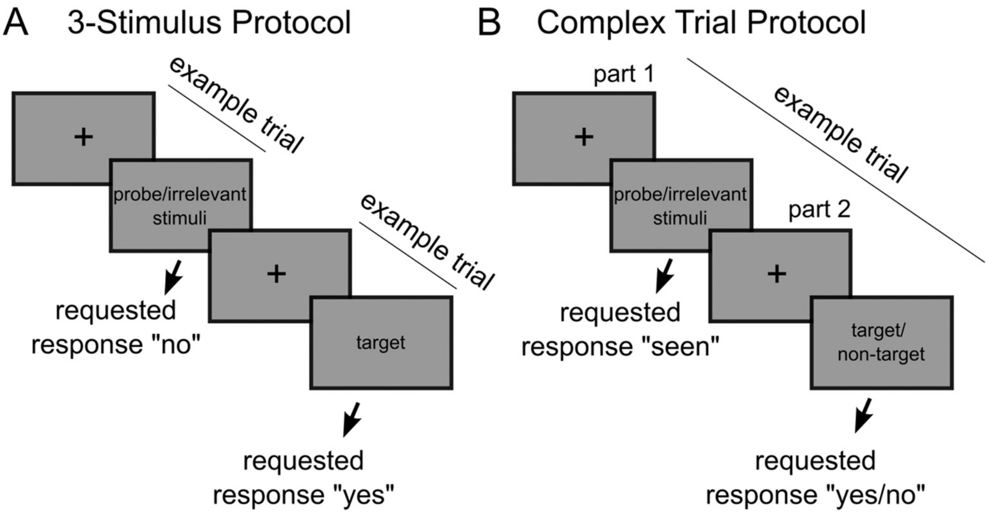


Fig. 2. Illustration of two trial protocols of the Concealed Information Test.
(A) In the "3-Stimulus Protocol", on each trial on of three stimuli categories are presented. The probe and irrelevant stimuli category is accompanied by a "target" category. Participants are instructed to respond "no" (not known/stimulus not recognized) to probe and irrelevant stimuli and "yes" to the target stimuli. **(B)** In the Complex Trial Protocol (CTP) each trial consists of two parts: In the first part, either a probe or an irrelevant stimulus is presented, and the participant must respond by pressing a key for "seen", regardless of the presented stimulus. In part two, either a target or non-target stimulus is presented, and the participant must decide whether the stimuli falls under the target or non-target category.

2.2.2. Complex Trial Protocol (CTP)

Given that over the years different CIT trial protocols have been used, it is important to consider trial protocols as a moderating factor. The originally P300 CIT protocol, the so-called “3-Stimulus Protocol” (Rosenfeld et al., 1988), includes three stimulus categories. Here, the probe and irrelevant stimulus categories are accompanied by a “target” category. Participants are instructed to respond “no” (not known/stimulus not recognized) to probe and irrelevant stimuli and “yes” to the target stimuli. According to Rosenfeld et al. (2006), the 3SP only generates submaximal P300 responses to probe stimuli because of the additional relevant stimuli category (target category), which leads to a reduction in the P300 response to the probe stimulus (Polich, 2007). In the Complex Trial Protocol (CTP) this issue is addressed by separating the decision about target vs. non-target and probe vs. irrelevant stimuli (Rosenfeld et al., 2018). Specifically, the CTP trial consists of two parts: In the first part, either a probe or an irrelevant stimulus is presented, and the participant must respond by pressing a key for “seen”, regardless of the presented stimulus. About one second later, either a target or non-target stimulus is presented, and the participant must decide whether the stimuli fall under the target or non-target category. The CTP has been applied to both the personal-item and mock-crime paradigm (Winograd and Rosenfeld, 2011; Winograd and Rosenfeld, 2014) and seems to have the highest accuracy when probe and irrelevant stimuli are presented as visual stimuli (Deng et al., 2016). The two trial protocols of the CIT are illustrated in Fig. 2.

2.2.3. Countermeasures

Another moderating factor is the application of countermeasures (CMs). These are behaviors used to manipulate physical parameters, such as motor movements of limbs (simple countermeasures, e.g., wiggling a toe of the left foot) or thoughts (mental countermeasures, e.g., thinking of something specific) during stimulus presentation to alter the response generated by the participant (Rosenfeld et al., 2008). Sasaki et al. (2001) showed that mental countermeasures (internally counting backward from seven) did not seem to have a significant effect on amplitude differences and thus the reliable assignment of guilty and innocent individuals, both Rosenfeld et al. (2004) and Mertens and Allen (2008) showed that the execution of simple countermeasures (e.g., pressing a finger when irrelevant stimuli are presented) could indeed be effective. They showed that irrelevant stimuli appeared more relevant in the EEG due to these measures, as they exhibited a higher amplitude and, therefore, could be less reliably distinguished from probe stimuli. CTP seems to be resistant to these countermeasures, although not immune (Rosenfeld et al., 2013; Rosenfeld and Labkovsky, 2010; Rosenfeld et al., 2008; Rosenfeld et al., 2017; Rosenfeld et al., 2018). Lukács et al. (2016) suggest that the repeatedly demonstrated resistance of CTP to these countermeasures could be explained by increased workload and could affect the amplitude, and thus, the secure assignment to the guilty or innocent group, which, in turn, directly affects reported effect sizes. CTP could, according to the current state of research, be more effective and reliable overall in comparison to 3SP when considering and taking countermeasures into account.

2.2.4. Inclusion of control participants

Meijer et al. (2007) argued that the inclusion of innocent participants as a control group can impact the effect size. In most P300 CIT studies, the control group (innocent condition) is omitted and the key analysis is to compare standardized responses to the relevant and neutral alternatives within subjects. As in the Meijer et al., 2014 meta-analysis we examined whether studies that included a control group resulted in different effect size estimates than studies that relied on within subject analyses.

2.2.5. Number of CIT questions (NQ)

Meijer et al. (2014) argue in their meta-analysis that both the number of Concealed Information Test (CIT) questions and the repetition of individual questions play a role in the reliability of a test according to psychometric theory. While the number of questions increases reliability because individual questions are independent of each other, the repetition of the same questions would lead to increased errors due to dependence (Meijer et al., 2014). The authors acknowledge the existing negative correlation between both factors, but they focus their analysis on the number of questions. This focus was motivated by a previous study by Ben-Shakhar and Elaad (2003), which showed that the number of CIT questions seems to be a very strong moderator. While this effect seems to be less pronounced in the context of the P300 (Meijer et al. (2014)), this number of questions will be considered as a moderator in the current study.

2.2.6. Year of publication

As with Meijer et al. (2014), this variable will also be controlled for, to take a possible decline in effect sizes over time into account. The decline effect is commonly observation in research areas and describes the observation that initial large effect sizes tend to decline with cumulation of research results (Schooler, 2011).

2.3. Signal detection measures d and a

In the context of this study, the signal detection measures Cohen's *d* and the *a*-value are used to quantify how well P300 amplitudes differences can distinguish between probe and distractor stimuli. Cohen's *d* is used as an effect size measure to estimate how much the population of guilty/knowledgeable individuals differs from the population of innocent/unaware individuals with respect to the amplitude of the P300 component in the CIT. Consequently, the primary outcome measure in this study are P300 amplitude differences elicited by probe versus distractor stimuli, which is assumed to reflect the differential recognition processes in guilty versus innocent individuals. Cohen's d was calculated according to the following formula:

$$d = \frac{\overline{X}_{probe} - \overline{X}_{control}}{s} \tag{1}$$

where $\overline{X}_{probe}$ is the mean amplitude of the probe stimulus, $\overline{X}_{control}$ is the mean amplitude of the irrelevant stimuli, and *s* is the pooled variance of the two stimulus categories.

The other measure that has been used to evaluate detection efficiency is the *a*-value (in some studies also denoted as A', (Zhang and Mueller, 2005)). This value represents the area under receiver operating characteristic (ROC) curve (AUC). The ROC curve is based on signal detection theory (Green and Swets, 1966). This method is used to estimate the degree of separation by a procedure between two groups or populations (Rosenfeld and Donchin, 2015).

These *d*- and *a*-values were directly extracted from the primary studies included in this meta-analysis. Each value was reported in the context of distinguishing between guilty and innocent individuals, either within the mock-crime paradigm or the personal-item paradigm. By relying on these reported measures, the current analysis ensures consistency with the original experimental designs and avoids recalculations that might overlook specific methodological nuances or subgroup effects considered in the original studies. Studies that did not provide *d*- or *a*-values explicitly were excluded from this analysis, ensuring that only standardized and comparable metrics were included. This approach aligns with the methodology proposed by Meijer et al. (2014) and ensures that the signal detection measures are representative of the detection efficiency as reported by the original authors.

Both values can be derived from each other, making them complementary, depending on which value is reported in a given study. The ROC curve and thus the *a*-value can be derived from Cohen's *d* using the Grier formula (Grier, 1971):

$$a = \phi\left(\frac{d}{\sqrt{2}}\right) \tag{2}$$

where ϕ represents the standard normal cumulated distribution function.

The non-parametric method employed in this study offers an advantage as it doesn't hinge on assumptions of normal distribution and equal variances. Although alternative methods, like bootstrapping of cross-correlations or Bayesian classification, were reported to potentially yield superior results, they were excluded from consideration due to the absence of standardized criteria for comparison across studies. It is noteworthy that all included studies quantified the P300 using the peak-to-peak method, as described by Mertens and Allen (2008). Here, a sliding window approach is employed to identify a 100-ms segment with the maximal positive amplitude average within the P300 latency window (300–800 ms). Following this, the method searches for a 100-ms segment with the greatest negative amplitude average in the period from 800 ms to 1300 ms. The difference between these two segments defines the P300 amplitude.

2.4. Simulation detection score distribution among innocent participants

Some CIT studies utilize both guilty and innocent participant groups, creating two experimental conditions that require a larger sample size for equivalent power compared to within-designs (Abraham and Russell, 2008). Consequently, an increasing number of studies only include guilty participants. Ben-Shakhar and Elaad (2003) considered such studies in their meta-analysis of the Autonomic Nervous System (ANS)-based CIT. In these cases, ROC areas were constructed by comparing distributions of probe and irrelevant stimuli within the guilty condition. However, this approach led to distorted estimates of d and a, contributing to a lack of external validity (Meijer et al., 2007), as the CIT aims to distinguish between guilty and innocent individuals, i.e., different individuals.

In many P300-based CIT studies, researchers rely on a simulated distribution of recognition values from bootstrap analyses, utilizing P300 amplitude differences between probe and irrelevant stimuli (Rosenfeld and Donchin, 2015). The procedure follows the following algorithm (Lu et al., 2018): The P300 amplitude of the probe stimulus is standardized within a person. This involves subtracting the mean P300 amplitude from all P300 amplitude values of used stimuli (e.g., for a probe and six irrelevant stimuli out of seven stimuli in total) and dividing by the standard deviation. This results in a standardized value for the probe stimulus, which should be greater than zero given the assumption that the P300 amplitude of the probe stimulus is larger in guilty individuals than that of irrelevant stimuli. This process is carried out for all individual participants in the guilty condition, creating standardized score distributions. To generate simulated innocent distributions, random values (as many as the number of stimuli, e.g., seven) are drawn from a standard normal distribution. One of these values serves as the probe value, as the probe stimulus in innocent individuals should have minimal impact. This P300 amplitude value of the probe is also standardized, as described earlier, and should be close to zero. This is repeated as many times as the number of guilty participants. Subsequently, ROCs and, consequently, a-values were computed from these actual and simulated values for their respective groups (Lu et al., 2018).

2.5. Correction of effect size d according to Schmidt and Hunter (2015)

As in the meta-analysis by Meijer et al. (2014), the study focused on the d-statistic, as its standard error is estimable (Schmidt and Hunter, 2015). Initially, the calculation and conversion from d to a and vice versa were performed, followed by the calculation of the corrected *d* values (d^*), detailed in Schmidt and Hunter (2015), the estimated standard error, and confidence intervals of d^* based on the formulas provided by Schmidt and Hunter (2015). The correction from *d* to d^* is based on a recommendation by Schmidt and Hunter (2015, p. 284–286), which suggests potential biases in estimators due to different sample sizes and was also conducted in the meta-analysis by Meijer et al. (2014). Specifically, corrected *d* values (d^*) were calculated for each study as follows:

$$d^* = \frac{d}{A}$$

where *A* is defined as (3)

$$A = 1 + \frac{0.75}{(Sample\ size - 3)}$$

The results of the meta-analysis are presented following the steps proposed by Harrer et al. (2021) and Döring and Bortz (2016), including the calculation of an overall effect size (pooled effect size), a heterogeneity analysis, examination of data asymmetry, potential assessment of publication bias, outlier and influence analysis, moderator analysis, and meta-regression analysis.

2.6. Random-effects model

To determine the mean effect size through the reported individual effect sizes of all included studies (SMD, standardized mean difference), a model was specified to weight individual studies. According to Döring and Bortz (2016), the fixed-effects model is only suitable when assuming that studies differ only due to sampling errors but capture the same population effect in terms of content. However, according to Harrer et al. (2021), this assumption does not reflect reality, as studies often differ from each other due to various factors such as different study conditions. Therefore, the validity of the fixed-effects model is questionable, as differences in certain treatment methods (or in this case, different paradigms or protocols) alone allow the possibility that one may be more effective than the other, and the potential heterogeneity between studies may not have arisen solely due to errors. The random-effects model, as suggested by Harrer et al. (2021), considers possible differences and systematic changes in effects. It weights individual studies based on the size of their estimated standard errors, so studies with smaller confidence intervals, i.e., more reliable estimates based on smaller estimated standard errors, receive higher weighting in the analysis (Harrer et al., 2021). Since a high level of heterogeneity is expected due to differences in conditions between studies, as mentioned earlier, the random-effects model was used for calculating an overall effect. Based on more robust results for an effect size relying on continuous outcome data, the Restricted Maximum Likelihood estimation (REML, Viechtbauer (2010)) was used. Additionally, the Knapp-Hartung adjustment (Knapp and Hartung, 2003) was applied, assuming a t-distribution of the data instead of a normal distribution, thereby reducing the risk of a false-positive result.

2.7. Software used and data availability

All analyses were performed in R (version 4.3.2; R Core Team, 2023) using a range of specialized packages. Meta-analyses and meta-regressions were conducted using the metafor package (Viechtbauer, 2010), while psychmeta (Dahlke and Wiernik, 2019) was used for Hunter-Schmidt corrections and moderator coding. meta (Balduzzi et al., 2019) and dmetar (Harrer et al., 2021) supported effect size estimation, subgroup analyses, and publication bias diagnostics (e.g., Egger's test, funnel plots). ggplot2 (Wickham, 2016) and sjPlot (Lüdecke, 2023) were used for data visualization, including interaction plots. readxl (Wickham and Bryan, 2025), readr (Wickham et al., 2024), and xlsx (Schauberger and Walker, 2025) were applied for data import, while dplyr (Wickham et al., 2025a), tidyr (Wickham et al., 2025b), and sjmisc (Lüdecke, 2018) supported data wrangling and transformation.

All code and data in this manuscript can be downloaded on the Open Science Framework at https://osf.io/yrzsc/.

3. Results

The results of the individual studies, the correction to d*, and information on the investigated moderators are presented in Table 2 (mock-crime paradigm) and Table 3 (personal-item paradigm). All subsequent analyses are based on *d** and the corrected estimated standard error and confidence intervals. The individual steps and their rationale are explained more specifically in the following subsections.

3.1. Mean effect size and heterogeneity analysis

The results can be found in the Forest plot in Fig. 3. The calculated overall effect size is d* = 1.59 and is significant with $t = 15.18$ ($p < 0.0001$, 95 % CI [1.38; 1.80]). The prediction interval of the pooled effect size ranges from CI [0.31, 2.87], indicating that the found effect sizes vary widely based on the available studies. This result suggests a smaller overall effect size d* than reported in the meta-analysis by Meijer et al. (2014), which reported d* = 1.89 for the P300, with a 95 % CI of [1.62; 2.15], which is slightly broader than that reported in this study.

After estimating the overall effect size, a test for heterogeneity among the included studies (Cochran's Q-Test) was conducted. The results of this analysis are shown in Table 4. The test was significant with $Q = 161.28$ ($p < 0.0001$), indicating that the assumption of homogeneity among the studies is rejected. To assess the extent of heterogeneity between studies, two measures were used. First, the restricted maximum likelihood estimator τ^2 (measure of inter-study variance) was employed, as the calculated effect sizes are based on continuous data (i.e., EEG measurements; Viechtbauer (2010)). This estimator is insensitive to the number of studies and their precision, providing an estimate of the variance of the true effect size based on the available studies. However, it cannot be compared across multiple studies. Second, the I^2 statistic was used to estimate the heterogeneity variance according to Higgins and Thompson (2002). This statistic is insensitive to the number of selected studies and defines the percentage of variability in effect sizes not caused by sampling error, providing an intuitively understandable measure. The heterogeneity variance between studies was $\tau^2 = 0.41$ (95 % CI [0.21; 0.70]), with an I^2 value of 67.8 % (95 % CI: [57.2 %; 75.7 %]), indicating moderate to substantial heterogeneity between the studies according to Higgins and Thompson (2002).

3.2. Moderator analysis

To better understand the heterogeneity in the data, we conducted moderator analysis (Harrer et al., 2021). Following the methodology of Schmidt and Hunter (2015) and Meijer et al. (2014), we only investigated moderator variables if the variance between studies exceeds the expected variance due to sampling error, indicating a rejection of the homogeneity assumption. Given the significant heterogeneity test with $Q = 161.28$ ($p < 0.001$) and the existing heterogeneity measures with $\tau^2 = 0.41$ and $I^2 = 67.8$ %, this condition was met. We utilized a mixed model for the analysis. This model combines a fixed-effect model and a random-effects model (Harrer et al., 2021). Specifically, this model assumes, akin to the random-effects model, that there is more than one true effect size, but these can be represented in subgroups, and within the subgroups, it is a fixed size. The results of the moderator analysis are presented in Table 5 and revealed that all variables, except for the CM, yield a moderating effect on the effect size. While CM variable provided significant effect sizes for each subgroup, the difference between subgroups in effect size is not significant, suggesting that this variable does not have a moderating effect.

In a control analysis we investigated whether the variable year of publication could be associated with the reported effect sizes. The analysis revealed a trend towards decreased effects sizes over the years ($\beta = -0.03$, $p = 0.08$).

3.3. Meta-regression

The estimated heterogeneity (I^2) in the first moderation analyses varied substantially, ranging from 57.2 % to 75.7 %. This suggests that there are significant differences between reported effect sizes, possibly due to other unconsidered variables and their expressions within each subgroup. To better understand this variability, we used a meta-regression approach introduced by Hedges and Pigott (2004) which

Table 2
P300 Results of the included Studies for mock-crime paradigm.

Study	Year	CTP	NQ	CM	Ng	Ni	CDR guilty	CDR innocent	BR	*d*	*d**	*a*
Ambach et al.*	2011	0	9	0	30	–	0.73	0.90	0.25	1.75	1.73	**0.892**
Funicelli et al.(Word, shallow vs. control)	2023	1	1	0	13	14			0.13	0.45	0.43	**0.624**
Funicelli et al. (Picture, shallow vs. control)	2023	1	1	0	14	14			0.13	0.98	0.95	**0.755**
Funicelli et al. (Word, deep vs. control)	2023	1	1	0	14	14			0.13	0.66	0.64	**0.679**
Funicelli et al. (Picture, deep vs. control)	2023	1	1	0	15	14			0.13	2.24	2.17	**0.943**
Hu and Rosenfeld(Immediate Condition)*	2012	1	1	0	12	12	**0.67**	**1.00**	0.11	1.73	1.67	**0.890**
Lu et al.	2018	1	1	0	18	–	**0.75**	0.90	0.16	1.41	1.36	**0.840**
Matsuda et al.*	2013	0	1	0	19	–	**0.79**	**0.84**	0.20	1.25	1.23	**0.812**
Meijer *	2008	0	6	0	30	–			0.17	**0.50**	0.49	**0.638**
Mertens & Allen (CM 0 Condition)*	2008	0	12	0	15	16	**0.47**	**1.00**	0.20	0.95	0.93	**0.750**
Mertens & Allen (CM 1 Condition)*	2008	0	12	1	15	16	**0.11**	**1.00**	0.20	0.04	0.03	**0.510**
Mertens & Allen (CM 2 Condition)*	2008	0	12	1	15	16	**0.20**	**1.00**	0.20	0.14	0.14	**0.540**
Mertens & Allen (CM 3 Condition)*	2008	0	12	1	15	16	**0.27**	**1.00**	0.20	0.29	0.28	**0.580**
Olson et al. (Episodic CTP – non-practice)	2022	1	1	1	18	–			0.14	0.55	0.53	**0.650**
Rosenfeld et al. Experiment 1, Week 1*	2004	1	6	0	11	11	**0.82**	**0.91**	0.20	2.40	2.28	0.955
Rosenfeld et al. Experiment 1, Week 2*	2004	0	6	1	11	11	**0.18**	**0.91**	0.20	0.55	0.52	**0.650**
Rosenfeld et al. (Experiment 1, add. subj)*	2004	0	6	0	11	–	**0.91**	**0.90**	0.20	2.29	2.20	**0.947**
Rosenfeld & Labkovsky (Guilty Condition)*	2010	1	1	0	12	13	**1.00**	**0.92**	0.14	2.09	2.02	0.930
Sai et al. (no-feedback Condition)	2020	1	1	0	18	18	**0.72**	**0.83**	0.16	0.74	0.73	**0.700**
Ward et al.	2020	1	1	0	19	–	**0.91**	**0.85**	0.16	0.91	0.89	**0.740**
Winograd & Rosenfeld (Guilty Condition)*	2011	1	1	0	12	12	**0.83**	**0.92**	0.14	2.09	2.02	0.930
Winograd & Rosenfeld (CM Condition)*	2011	1	1	1	12	12	**1.00**	**0.92**	0.14	2.90	2.80	**0.980**
Winograd & Rosenfeld (Innocent-naive/Guilty-naive)	2014	1	1	0	14	14	**0.79**	**0.86**	0.16	1.48	1.43	**0.852**

Note: NQ = Number of CIT questions; Ng = Number of knowledgeable (guilty) participants; CDR guilty = Correct Detection Rate among knowledgeable participants; Ni = Number of unknowledgeable (innocent) participants; CDR or expected CDR innocent = Correct Detection Rate among unknowledgeable participants; BR = Proportion of relevant items within each question. d = Effect size measure; a = Area under the ROC curve. Bolded numbers indicate that the corresponding measure was directly taken from the article. If none of the numbers are bold they are calculated by other given measures (e.g. mean and SD). Studies marked with * were also included in the two most recent meta-analyses on the CIT and P300 (Meijer et al. (2014) and Leue and Beauducel (2019)).

Table 3
P300 Results of the included studies for personal-item paradigm.

Study	Year	CTP	NQ	CM	Ng	Ni	CDR guilty	CDR innocent	BR	*d*	*d**	*a*
Funicelli et al. (SG vs. IC)	2021	1	1	0	12	12			0,14	1,71	1.65	0.887
Funicelli et al. (GCM vs. IC)	2021	1	1	1	12	12			0,14	1,19	1.15	0.800
Hsu et al.	2024	0	1	1	41	–			0,17	1,31	1.30	0.823
Hu et al. (Guilty Condition)*	2012	1	1	0	12	12	**1.00**	**0.92**	0.11	2.88	2.78	**0.979**
Hu et al. (CM 2/6 Condition)*	2012	1	1	1	13	12	**0.92**	**0.92**	0.11	2.48	2.39	**0.960**
Hu et al. (CM 4/6 Condition)*	2012	1	1	1	12	12	**0.83**	**0.92**	0.11	2.09	2.02	**0.930**
Hu et al. (CM 6/6 Condition)*	2012	1	1	1	14	12	**0.71**	**0.92**	0.11	1.73	1.68	**0.890**
Lukács, Weiss et al. (Control/Guilty)	2016	1	1	0	15	14	**0.93**	**0.93**	0.14	2.80	2.72	**0.980**
Lukács, Weiss et al. (Control/Old CM)	2016	1	1	1	16	14	**0.80**	**0.93**	0.14	2.09	2.03	**0.930**
Lukács, Weiss et al. (Control/New CM)	2016	1	1	1	15	14	**0.88**	**0.93**	0.14	2.20	2.13	**0.940**
Meijer et al.*	2007	0	1	0	24	–	**0.92**	**0.95**	0.20	2.54	2.50	0.964
Meixner et al. (Experiment 1)*	2009	1	1	0	12	12	**1.00**	**0.80**	0.19	2.44	2.36	0.958
Meixner et al. (Experiment 2)*	2009	1	1	0	15	15	**0.93**	**0.80**	0.19	2.05	1.99	0.926
Meixner & Rosenfeld*	2011	1	1	0	12	–	**0.83**	0.90	0.19	2.03	1.96	0.924
Olson et al. (Semantic CTP – non-practice)	2022	1	1	1	23	–			0.14	1.66	1.63	**0.880**
Rosenfeld et al. Experiment 2, Week 1 Guilty*	2004	0	1	0	13	–	**0.92**	**0.90**	0.20	2.34	2.27	**0.951**
Rosenfeld et al. Experiment 2, Week 1 Control*	2004	0	1	0	10	–	**0.90**	**0.90**	0.20	2.25	2.15	**0.944**
Rosenfeld et al., Target Condition *	2006	0	1	0	10	–	**0.90**	0.90	0.20	2.25	2.15	0.944
Rosenfeld et al., No Target Condition*	2006	0	1	0	11	–	**0.82**	0.90	0.17	1.98	1.90	0.919
Rosenfeld et al. (Auto 1 PB)*	2007	0	3	0	13	–	**0.62**	0.90	0.20	1.48	1.43	0.853
Rosenfeld et al. (Auto 6 PB)*	2007	0	6	0	9	–	**0.56**	0.90	0.20	1.36	1.30	0.832
Rosenfeld et al. Experiment 1 Week 1*	2008	1	1	0	12	12	**0.92**	**0.92**	0.18	2.4	2.28	0.955
Rosenfeld et al. (Symmetric)*	2009	1	1	0	12	12	**0.83**	**0.92**	0.19	2.88	2.78	0.979
Rosenfeld et al. (Asymmetric)*	2009	1	1	0	12	–	**1.00**	0.90	0.19	2.77	2.68	0.975
Rosenfeld & Labkovsky (Guilty Condition)*	2010	1	1	0	13	13	**1.00**	**0.92**	0.19	2.93	2.84	0.981
Rosenfeld & Labkovsky (CM Condition)*	2010	1	1	1	12	13	**1.00**	**0.92**	0.19	2.48	2.39	**0.960**
Rosenfeld et al. (Deception Condition)*	2012	0	1	0	10	–	**1.00**	0.90	0.14	2.77	2.65	0.975
Rosenfeld et al. (Control Condition)*	2012	0	1	0	10	–	**0.50**	0.90	0.14	1.25	1.19	0.811
Sokolovsky et al.*	2011	1	1	0	12	–	**0.83**	0.90	0.19	2.03	1.94	0.924
Verschuere, Rosenfeld et al. (Deception)*	2011	0	1	0	16	–	**0.75**	0.90	0.20	1.79	1.74	0.897
Verschuere, Rosenfeld et al. (Truth)*	2011	0	1	0	18	–	**0.44**	0.90	0.20	1.14	1.11	0.789

Note: NQ = Number of CIT questions; Ng = Number of knowledgeable (guilty) participants; CDR guilty = Correct Detection Rate among knowledgeable participants; Ni = Number of unknowledgeable (innocent) participants; CDR or expected CDR innocent = Correct Detection Rate among unknowledgeable participants; BR = Proportion of relevant items within each question. d = Effect size measure; a = Area under the ROC curve. Bolded numbers indicate that the corresponding measure was directly taken from the article. If none of the numbers are bold they are calculated by other given measures (e.g. mean and SD). Studies marked with * were also included in the two most recent meta-analyses on the CIT and P300 (Meijer et al. (2014) and Leue and Beauducel (2019)).

allowed us to study interaction effects between the moderators in an exploratory analysis.

To avoid overfitting (Gigerenzer, 2004) we used the most parsimonious model for our analyses. For model comparison, all significant moderators from the moderation analyses (paradigm, CTP, CM (countermeasures), NQ (number of questions), and CG (control group)) were considered. The best model was initially determined through a multimodel inference analysis, considering possible interactions (Harrer et al., 2021). This method tests all possible predictor combinations and their relevance exploratively based on the available data. The best model included the variables paradigm, CTP and CM (AIC = 101.1, BIC = 117.1). The inclusion of the predictors CG or NQ did not result in a better model fit (AIC = 101.7, BIC = 127.6, and AIC = 104.6, BIC = 124.5, respectively). The results were further verified through hierarchical regression. Both procedures showed the same result: including the variables CG and NQ did not significantly improve the regression model concerning the explanation of heterogeneity between studies. This may suggest that, considering other influences, these variables cannot explain a significant change in effect size. Furthermore, it replicates results by Meijer et al. (2014) that both the number of questions and the use of control groups do not significantly contribute to the change in effect size and, thus, the accuracy of distinguishing between guilty and innocent individuals.

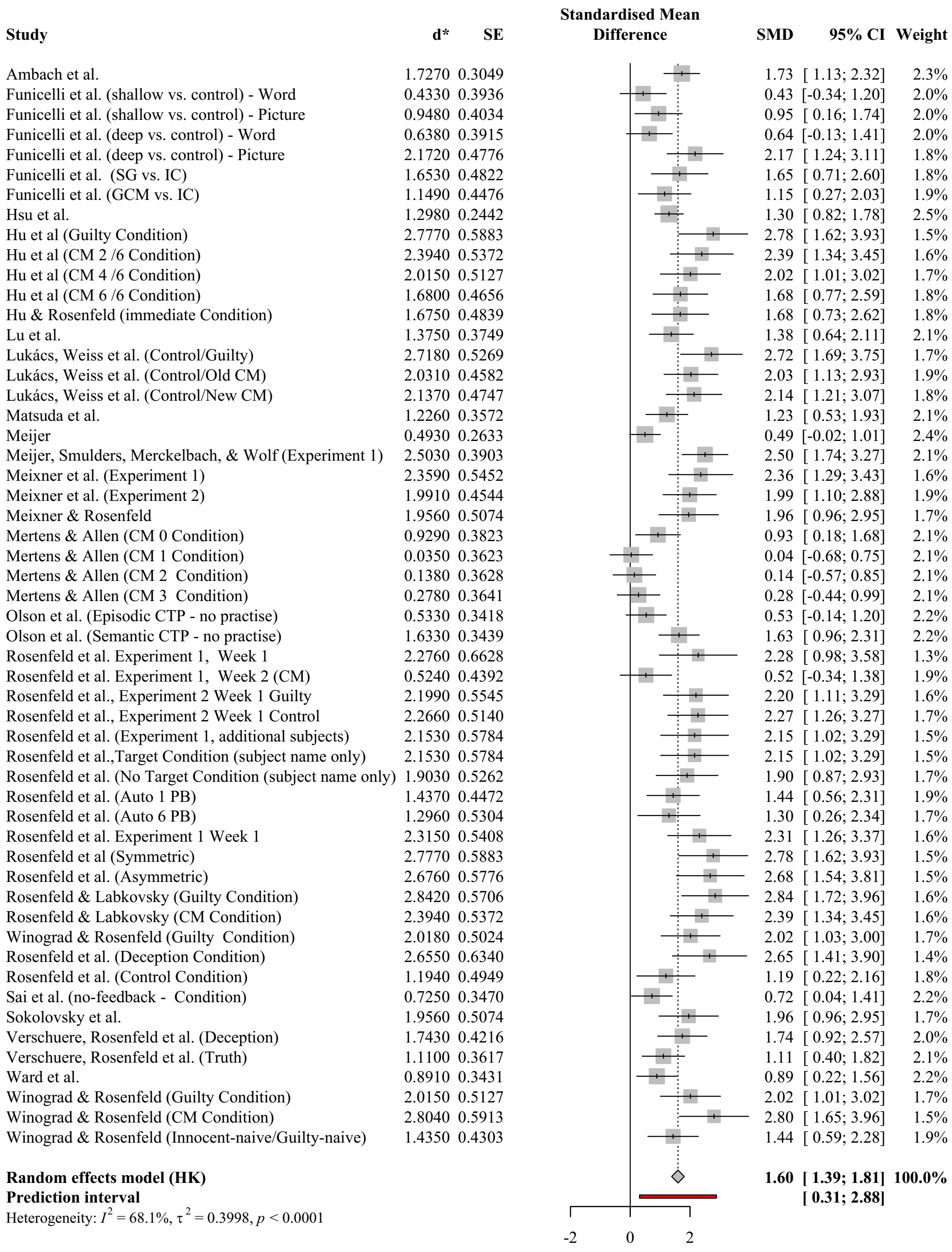
The model with the best fit for the available data was then subjected to a permutation test to better assess the robustness of the predictors (Higgins and Thompson, 2002). The significant results from the meta-regression model remained robust in this test. We found that the model with the predictors paradigm (x_1), CTP (x_1), and CM (x_1) and their interactions predicted the differences in effect sizes best. The results of the multiple meta-regression analysis are presented in Table 6.

The mixed-model statistics indicated a residual heterogeneity (τ^2) of 0.12, with a corresponding τ of 0.34. The I^2 value was significantly reduced to 37 %, suggesting lower unaccounted variability. The R^2 value, indicating the proportion of variance explained by the model, was high at 70 %. The ANOVA test for the moderators, which assessed the overall significance of the included moderators, was significant (F (6, 47) = 9.02, p < 0.001), confirming that the moderators collectively explain a substantial portion of the heterogeneity.

The results of the meta-regression revealed that considering individual variables explains R^2 = 70 % of the heterogeneity between studies. Furthermore, it highlights that a simple moderator analysis, i.e., an analysis of variables without considering the influence of other variables, may lead to incorrect conclusions. The previously reported heterogeneity between studies of I^2 = 67.8 % (indicating heterogeneity not caused by sampling error) decreases to τ^2 = 0.12 and I^2 = 37 % by considering additional predictors, which, according to Higgins and Thompson (2002), corresponds to lower heterogeneity.

The meta-regression revealed only a significant main effect for the variable countermeasures (CM; $\beta = -1.05$, $p < 0.01$). No other main effect emerged (see Table 4). This result is at odds with the results of the meta-analysis by Meijer et al. (2014), which demonstrated that the choice of the personal-item paradigm positively affects the effect size and the reliable identification of guilty individuals. Our results suggest that, the choice of the personal-item paradigm alone does not lead to a significantly larger effect size, whereas the use of CM methods results in a significantly smaller effect size (d*).

Interestingly, the moderator analysis described above indicated a significant moderating effect for the variable CTP. However, the meta-regression, which considered interaction effects, showed no significant main effect for the protocol used (CTP; $\beta = -0.06$, $p = 0.83$). Interaction



Study	d*	SE	SMD	95% CI	Weight
Ambach et al.	1.7270	0.3049	1.73	[1.13; 2.32]	2.3%
Funicelli et al. (shallow vs. control) - Word	0.4330	0.3936	0.43	[-0.34; 1.20]	2.0%
Funicelli et al. (shallow vs. control) - Picture	0.9480	0.4034	0.95	[0.16; 1.74]	2.0%
Funicelli et al. (deep vs. control) - Word	0.6380	0.3915	0.64	[-0.13; 1.41]	2.0%
Funicelli et al. (deep vs. control) - Picture	2.1720	0.4776	2.17	[1.24; 3.11]	1.8%
Funicelli et al. (SG vs. IC)	1.6530	0.4822	1.65	[0.71; 2.60]	1.8%
Funicelli et al. (GCM vs. IC)	1.1490	0.4476	1.15	[0.27; 2.03]	1.9%
Hsu et al.	1.2980	0.2442	1.30	[0.82; 1.78]	2.5%
Hu et al (Guilty Condition)	2.7770	0.5883	2.78	[1.62; 3.93]	1.5%
Hu et al (CM 2 /6 Condition)	2.3940	0.5372	2.39	[1.34; 3.45]	1.6%
Hu et al (CM 4 /6 Condition)	2.0150	0.5127	2.02	[1.01; 3.02]	1.7%
Hu et al (CM 6 /6 Condition)	1.6800	0.4656	1.68	[0.77; 2.59]	1.8%
Hu & Rosenfeld (immediate Condition)	1.6750	0.4839	1.68	[0.73; 2.62]	1.8%
Lu et al.	1.3750	0.3749	1.38	[0.64; 2.11]	2.1%
Lukács, Weiss et al. (Control/Guilty)	2.7180	0.5269	2.72	[1.69; 3.75]	1.7%
Lukács, Weiss et al. (Control/Old CM)	2.0310	0.4582	2.03	[1.13; 2.93]	1.9%
Lukács, Weiss et al. (Control/New CM)	2.1370	0.4747	2.14	[1.21; 3.07]	1.8%
Matsuda et al.	1.2260	0.3572	1.23	[0.53; 1.93]	2.1%
Meijer	0.4930	0.2633	0.49	[-0.02; 1.01]	2.4%
Meijer, Smulders, Merckelbach, & Wolf (Experiment 1)	2.5030	0.3903	2.50	[1.74; 3.27]	2.1%
Meixner et al. (Experiment 1)	2.3590	0.5452	2.36	[1.29; 3.43]	1.6%
Meixner et al. (Experiment 2)	1.9910	0.4544	1.99	[1.10; 2.88]	1.9%
Meixner & Rosenfeld	1.9560	0.5074	1.96	[0.96; 2.95]	1.7%
Mertens & Allen (CM 0 Condition)	0.9290	0.3823	0.93	[0.18; 1.68]	2.1%
Mertens & Allen (CM 1 Condition)	0.0350	0.3623	0.04	[-0.68; 0.75]	2.1%
Mertens & Allen (CM 2 Condition)	0.1380	0.3628	0.14	[-0.57; 0.85]	2.1%
Mertens & Allen (CM 3 Condition)	0.2780	0.3641	0.28	[-0.44; 0.99]	2.1%
Olson et al. (Episodic CTP - no practise)	0.5330	0.3418	0.53	[-0.14; 1.20]	2.2%
Olson et al. (Semantic CTP - no practise)	1.6330	0.3439	1.63	[0.96; 2.31]	2.2%
Rosenfeld et al. Experiment 1, Week 1	2.2760	0.6628	2.28	[0.98; 3.58]	1.3%
Rosenfeld et al. Experiment 1, Week 2 (CM)	0.5240	0.4392	0.52	[-0.34; 1.38]	1.9%
Rosenfeld et al., Experiment 2 Week 1 Guilty	2.1990	0.5545	2.20	[1.11; 3.29]	1.6%
Rosenfeld et al., Experiment 2 Week 1 Control	2.2660	0.5140	2.27	[1.26; 3.27]	1.7%
Rosenfeld et al. (Experiment 1, additional subjects)	2.1530	0.5784	2.15	[1.02; 3.29]	1.5%
Rosenfeld et al.,Target Condition (subject name only)	2.1530	0.5784	2.15	[1.02; 3.29]	1.5%
Rosenfeld et al. (No Target Condition (subject name only)	1.9030	0.5262	1.90	[0.87; 2.93]	1.7%
Rosenfeld et al. (Auto 1 PB)	1.4370	0.4472	1.44	[0.56; 2.31]	1.9%
Rosenfeld et al. (Auto 6 PB)	1.2960	0.5304	1.30	[0.26; 2.34]	1.7%
Rosenfeld et al. Experiment 1 Week 1	2.3150	0.5408	2.31	[1.26; 3.37]	1.6%
Rosenfeld et al (Symmetric)	2.7770	0.5883	2.78	[1.62; 3.93]	1.5%
Rosenfeld et al. (Asymmetric)	2.6760	0.5776	2.68	[1.54; 3.81]	1.5%
Rosenfeld & Labkovsky (Guilty Condition)	2.8420	0.5706	2.84	[1.72; 3.96]	1.6%
Rosenfeld & Labkovsky (CM Condition)	2.3940	0.5372	2.39	[1.34; 3.45]	1.6%
Winograd & Rosenfeld (Guilty Condition)	2.0180	0.5024	2.02	[1.03; 3.00]	1.7%
Rosenfeld et al. (Deception Condition)	2.6550	0.6340	2.65	[1.41; 3.90]	1.4%
Rosenfeld et al. (Control Condition)	1.1940	0.4949	1.19	[0.22; 2.16]	1.8%
Sai et al. (no-feedback - Condition)	0.7250	0.3470	0.72	[0.04; 1.41]	2.2%
Sokolovsky et al.	1.9560	0.5074	1.96	[0.96; 2.95]	1.7%
Verschuere, Rosenfeld et al. (Deception)	1.7430	0.4216	1.74	[0.92; 2.57]	2.0%
Verschuere, Rosenfeld et al. (Truth)	1.1100	0.3617	1.11	[0.40; 1.82]	2.1%
Ward et al.	0.8910	0.3431	0.89	[0.22; 1.56]	2.2%
Winograd & Rosenfeld (Guilty Condition)	2.0150	0.5127	2.02	[1.01; 3.02]	1.7%
Winograd & Rosenfeld (CM Condition)	2.8040	0.5913	2.80	[1.65; 3.96]	1.5%
Winograd & Rosenfeld (Innocent-naive/Guilty-naive)	1.4350	0.4303	1.44	[0.59; 2.28]	1.9%
Random effects model (HK)			**1.60**	**[1.39; 1.81]**	**100.0%**
Prediction interval				**[0.31; 2.88]**	

Heterogeneity: $I^2 = 68.1\%$, $\tau^2 = 0.3998$, $p < 0.0001$

Fig. 3. Results of the calculation of the overall effect size and weighting of individual studies in the meta-analysis. Weighting is expressed as a percentage, d^* = corrected effect size of individual studies, Number of combined studies: $k = 54$, SMD = model-(rounded) values for Effect size d* (standardized mean difference).

Table 4
Results of the test for heterogeneity and measures of heterogeneity for heterogeneity analysis.

Test for heterogeneity	Q	df	p - value
	161.28	52	<0.0001

Measures of heterogeneity	τ^2	95 % - CI of τ^2	I^2	95 % - CI of I^2
	0.41	[0.21; 0.70]	67.8 %	[57.2 % ; 75.7 %]

effects of the meta-regression demonstrated that CTP is less affected by counter measures ($\beta = 1.11$, $p < 0.05$; Fig. 4A). Moreover, there was a trend towards a paradigm × CTP interaction ($\beta = 0.66$, $p = 0.08$). Here, the combination of the personal item and CTP may lead to increased effect sizes (see Fig. 4B).

Together, the results of the meta-regression advocate for a widespread use of CTP, particularly in scenarios involving countermeasures or when the personal-item paradigm is applied. By employing CTP, practitioners can enhance the reliability and diagnostic accuracy of the P300-based Concealed Information Test (CIT).

Table 5
Results of the moderator analysis.

	*d**	95 % - CI	*p* -value	I^2	95 % - CI	*p* – value for subgroup
Paradigm						<0.001
Mock-Crime	1.11	[0.76; 1.45]	<0.001	0.67	[0.49; 0.79]	
Personal-Item	1.95	[1.75; 2.14]	<0.001	0.21	[0.00; 0.49]	
CTP						0.003
Yes	1.76	[1.50; 2.02]	0.001	0.60	[0.41; 0.73]	
No	1.35	[1.00; 1.70]	<0.001	0.71	[0.55; 0.81]	
CM						0.290
Yes	1.35	[0.80; 1.89]	<0.001	0.78	[0.64; 0.87]	
No	1.67	[1.45; 1.89]	<0.001	0.59	[0.42; 0.71]	
NQ						<0.001
<5	1.76	[1.55; 1.97]	<0.001	0.55	[0.36; 0.68]	
>5	0.89	[0.33; 1.46]	0.005	0.71	[0.45; 0.85]	
CG						<0.001
Yes	1.61	[1.29; 1.94]	<0.001	0.73	[0.61; 0.81]	
No	1.56	[1.30; 1.94]	<0.001	0.58	[0.31; 0.73]	

Note: CTP = Complex Trial Protocol, CM = Countermeasures, NQ = Number of Questions, CG = Control Group. This table presents the effect sizes (*d**), 95 % confidence intervals (CI), *p*-values, and *F*-values for the variables Paradigm, CTP, Countermeasures (CM), Number of Questions (NQ), and Control Group (CG) in the context of P300 based-Concealed Information Tests (CIT). Significant differences are observed across the variables, indicating their influence on the effect sizes.

Table 6
Results of the meta-regression. The mixed-model statistics show a low residual heterogeneity (τ^2 = 0.12; I^2 = 37 %), indicating that the model successfully accounts for most of the variance between studies. The model's goodness-of-fit is further supported by a high R^2 value of 70 %. The ANOVA test for the moderators (F(6, 47) = 9.02, p < 0.001) confirms the significant contribution of the moderators to explaining heterogeneity.

Model-results	*β*	SE	t-value	*p*-Value	95 % - CI
Intercept	1.30	0.24	5.50	<0.001	[0.81; 1.77]
Main effects					
Paradigm	0.45	0.27	1.70	0.10	[−0.08; 0.99]
CTP	−0.06	0.27	−0.22	0.83	[−0.59; 0.48]
CM	−1.05	0.33	−3.19	<0.01	[−1.76; −0.36]
Interaction effects					
Paradigm × CTP	0.66	0.38	1.73	0.09	[−0.08; 1.38]
Paradigm × CM	−0.50	0.51	−0.99	0.34	[−1.52; 0.52]
CTP × CM	1.11	0.54	2.08	<0.05	[0.03; 2.19]

Note: CTP = Complex Trial Protocol, CM = Countermeasures, NQ = Number of Questions, CG = Control Group.

4. Discussion

In the current meta-analysis, we aimed to offer a comprehensive and up-to-date synthesis of the literature on the P300-based Concealed Information Test (CIT), evaluating its overall effectiveness and identifying critical moderating factors that influence its diagnostic accuracy. Key results are summarized in Fig. 5. We observed a large and statistically significant overall effect size, aligning with prior meta-analyses and supporting the use of the P300 as a reliable marker for detecting concealed knowledge. However, the presence of moderate to high heterogeneity across studies warranted further examination through moderator analyses and meta-regression.

Our moderator analysis indicated that, when considered individually, variables such as paradigm, trial protocol (3-stimulus vs. complex protocols), number of CIT questions, and the inclusion of control groups were each associated with variability in effect sizes. Only the use of countermeasures (CM) did not show a significant simple moderation effect. These findings differ from those reported by Meijer et al. (2014), who identified significant effects only for the paradigm variable. This discrepancy likely reflects the inclusion of more recent studies and greater methodological variability in the present dataset.

To assess the independent contributions of these moderators—and to test for possible interaction effects—we applied a meta-regression approach. This analysis identified three important moderators: paradigm, protocol, and countermeasures. Interestingly, the main effect of the paradigm variable (personal-item vs. mock-crime) was not significant in the meta-regression, despite its significance in the individual moderator analysis. This suggests that the previously reported superiority of personal-item paradigms (e.g., Meijer et al., 2014) may be less robust when accounting for interdependence of moderators. Similarly, CTP alone did not emerge as a significant predictor in the full model. However, its interaction with countermeasures indicates that CTPs help preserve effect sizes when countermeasure techniques are employed. Therefore, CTPs should be preferred over 3-stimulus protocols, particularly in contexts where the use of countermeasures cannot be ruled out.

In terms of cognitive processes engaged during deception tasks, our findings align with those of Leue and Beauducel (2019), who emphasized the cognitive complexity underlying P300 responses in deception paradigms. Elevated P300 amplitudes in response to concealed stimuli likely reflect their increased salience (e.g., Kok, 2001) as well as the mental effort involved in the deception process (Beauducel et al., 2006). However, additional cognitive processes—such as orienting responses, memory activation, and emotional or motivational factors—may also contribute to P300 differences observed in CITs paradigms (e.g., Verschuere and Ben-Shakhar, 2011). These processes are likely superimposed within the P300 signal. Future studies should aim to disentangle these components, for example through principal component analysis. The significant effect of countermeasures suggests that the use of physical and mental countermeasure techniques towards irrelevant stimuli reduces the P300 difference between irrelevant and concealed items—presumably by increasing P300 amplitudes to irrelevant stimuli. Importantly, the observed interaction between countermeasures and trial protocol indicates that this negative effect can be mitigated through the use of the CTP.

In terms of practical implications, our findings support the forensic utility of the P300-based CIT, particularly when implemented using the CTP. Although field applications have historically been limited, recent advances in mobile EEG technology (Biondi et al., 2022) offer promising opportunities to extend the CIT beyond laboratory settings. This will be a crucial next step in validating its real-world applicability for several reasons.

First, participant samples in the included studies were highly homogeneous. Most involved healthy, young adults—typically university students without psychiatric or neurological diagnoses. In contrast, real forensic populations often include individuals with substance use disorders, depression, or cognitive impairments (Baillargeon et al., 2009; Fazel and Danesh, 2002). These conditions are known to alter P300 characteristics, including amplitude and latency (Polich and Herbst, 2000; Porjesz and Begleiter, 2003), and are associated with reduced P300 responses to emotional or personally relevant stimuli (Bruder et al., 2011; Ford and Mathalon, 2004). Therefore, testing the CIT in more representative samples is essential.

Second, all studies included in the meta-analysis were conducted under controlled laboratory conditions. While this allows for precise experimental manipulation, it limits ecological validity. The absence of field studies—where suspects are evaluated under real investigative circumstances—reduces the generalizability of the findings. Despite the inherent challenges of establishing ground truth in applied settings (Meijer et al., 2014), such studies are critical for assessing whether the CIT performs reliably under conditions of social stress, high stakes, and environmental variability.

4.1. Limitations

This meta-analysis provides a comprehensive overview of the current

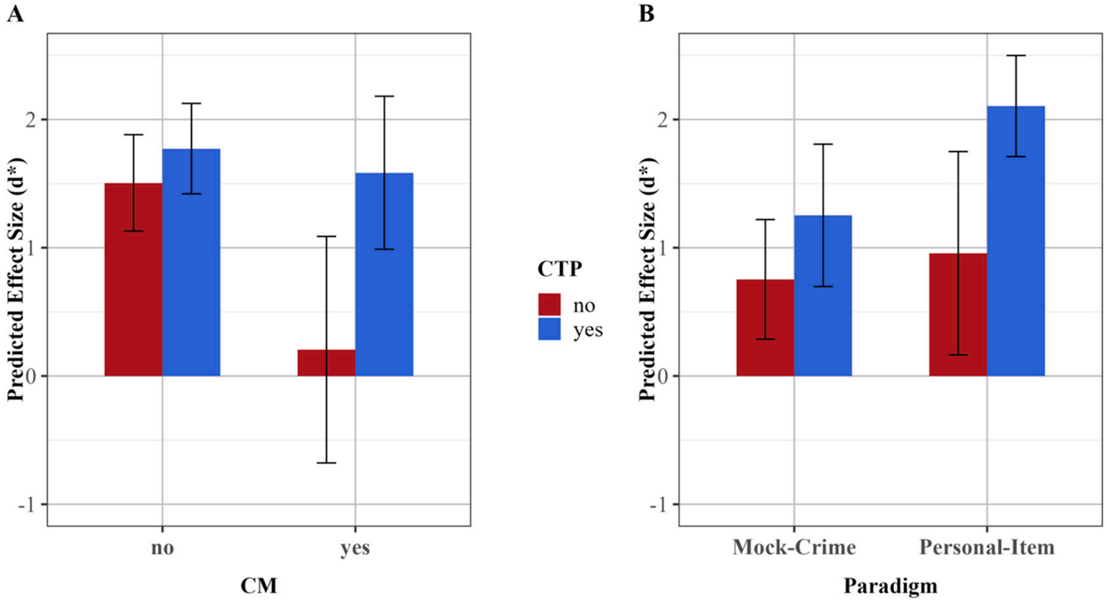
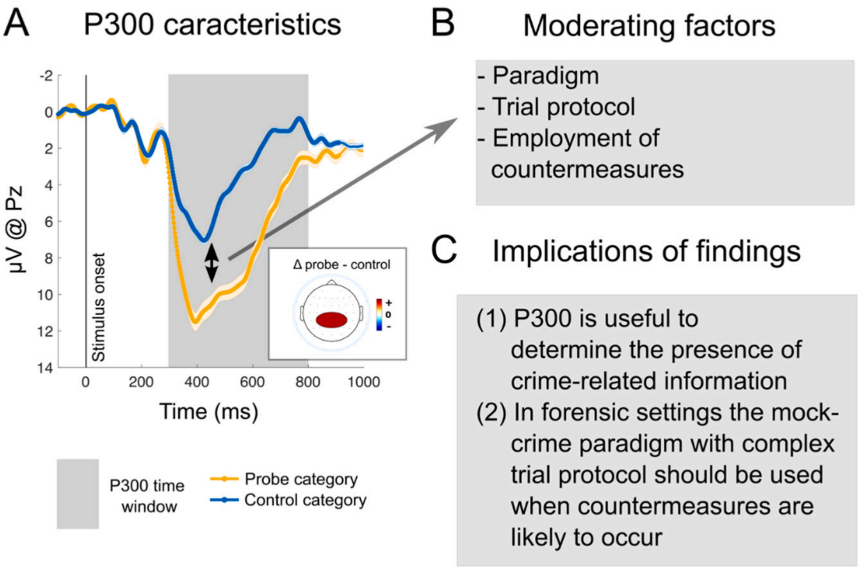


Fig. 4. The Complex Trial Protocol (CTP) should be used when countermeasures or personal-item paradigm is applied. This figure shows the predicted effect-size d* based on the interaction of the chosen paradigm, countermeasures (CM), and the use of the CTP protocol. The figure illustrates that the CTP protocol consistently improves effect sizes, particularly in the presence of countermeasures **(A)** and when the personal-item paradigm is applied **(B).** *Note:* Error bars reflect 95 % confidence intervals.



B Moderating factors

- Paradigm
- Trial protocol
- Employment of countermeasures

C Implications of findings

(1) P300 is useful to determine the presence of crime-related information
(2) In forensic settings the mock-crime paradigm with complex trial protocol should be used when countermeasures are likely to occur

Fig. 5. Characteristics of primary outcome measure and summary of key findings.
(A) Typical P300 responses elicited by relevant (probe) and control alternative in a CIT. Larger (i.e., more positive) amplitudes in the P300 time window (300–800 ms) are observed after the presentation of probe stimuli as compared to control stimuli. The inset shows the centro-parietal topographical map of this pattern of results. **(B)** Our moderation analysis shows that P300 effect in CIT are affected by the choice of paradigm (personal item vs. mock-crime paradigm), the chosen trial protocol (complex vs. original) and the likelihood of subjects to employ countermeasures. **(C)** Based on our findings, we conclude that the P300 is useful to determine the presence of crime-related information and that people interesting in using the CIT should use the mock-crime paradigm with the complex trial protocol.

state of research on the P300-based Concealed Information Test. However, several limitations should be noted.

First, the inclusion of studies was limited to peer-reviewed journal articles published in English. This introduces the possibility of publication bias, as studies reporting null or negative results are less likely to be published. Although we followed a systematic search strategy and conducted backward reference screening, the absence of a formal assessment of publication bias (e.g., funnel plot, trim-and-fill) may have left potential bias unaccounted for.

Second, although several relevant moderators were included in this analysis, some individual difference variables that may influence P300 responses, such as gender, emotional state, or trait anxiety, could not be examined due to inconsistent reporting. Previous research has suggested that these variables may interact with task demands and affect the neural correlates of recognition and deception, but the evidence remains too limited for formal inclusion in meta-analytic models.

Third, the analysis focused solely on P300 amplitude as the primary dependent variable. Other important features, such as peak latency, intra-individual variability, or topographic distribution, were not consistently reported and therefore excluded. These metrics could provide additional insights into the cognitive mechanisms underlying concealed information detection (Duncan et al., 2009; Polich, 2007).

4.2. Summary and conclusions

This meta-analysis presents a detailed synthesis on the diagnostic validity of the P300-based Concealed Information Test (CIT). Expanding upon prior work (e.g., Meijer et al., 2014; Leue and Beauducel, 2019), we included 16 additional studies and examined a broader range of potential moderators using advanced statistical models. Our findings suggest that, under controlled laboratory conditions, the P300-based CIT is a highly valid technique for detecting concealed information. The Complex Trial Protocol (CTP) should be preferred over standard 3-stimulus protocols, particularly in contexts where the use of counter-measures cannot be ruled out.

Despite robust laboratory-based evidence, the translation of the CIT into applied forensic settings remains limited. Technological constraints and the need for trained personnel have historically posed barriers to implementation. However, recent advances in mobile EEG systems and protocol optimization offer promising avenues for broader real-world application.

In conclusion, this meta-analysis reaffirms the diagnostic value of the P300-based CIT and provides critical guidance for maximizing its effectiveness. With continued refinement, validation in more diverse populations, and the integration of portable EEG technology, the CIT holds strong potential as a scalable, objective tool for forensic investigations.

CRediT authorship contribution statement

Julia Knappe: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Markus Ullsperger:** Writing – review & editing, Supervision, Resources, Conceptualization. **Hans Kirschner:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpsycho.2025.113236.

Data availability

All code and data in this manuscript can be downloaded on the Open Science Framework at https://osf.io/yrzsc/.

References

Abraham, W.T., Russell, D.W., 2008. Statistical power analysis in psychological research. Soc. Personal. Psychol. Compass 2 (1), 283–301. https://doi.org/10.1111/j.1751-9004.2007.00052.x.

Ambach, W., Dummel, S., Lüer, T., Vaitl, D., 2011. Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. Int. J. Psychophysiol. 81 (3), 275–282. https://doi.org/10.1016/j.ijpsycho.2011.07.010.

Baillargeon, J., Binswanger, I.A., Penn, J.V., Williams, B.A., Murray, O.J., 2009. Psychiatric disorders and repeat incarcerations: the revolving prison door. Am. J. Psychiatry 166 (1), 103–109. https://doi.org/10.1176/appi.ajp.2008.08030416.

Balduzzi, S., Rücker, G., Schwarzer, G., 2019. How to perform a meta-analysis with R: a practical tutorial. Evid. Based Ment. Health 22 (4), 153–160. https://doi.org/10.1136/ebmental-2019-300117.

Beauducel, A., Brocke, B., Leue, A., 2006. Energetical bases of extraversion: effort, arousal, EEG, and performance. Int. J. Psychophysiol. 62 (2), 212–223. https://doi.org/10.1016/j.ijpsycho.2005.12.001.

Ben-Shakhar, G., Elaad, E., 2003. The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. J. Appl. Psychol. 88 (1), 131–151. https://doi.org/10.1037/0021-9010.88.1.131.

Biondi, A., Santoro, V., Viana, P.F., Laiou, P., Pal, D.K., Bruno, E., Richardson, M.P., 2022. Noninvasive mobile EEG as a tool for seizure monitoring and management: a systematic review. Epilepsia 63 (5), 1041–1063. https://doi.org/10.1111/epi.17220.

Bond Jr., C.F., DePaulo, B.M., 2006. Accuracy of deception judgments. Personal. Soc. Psychol. Rev. 10 (3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2.

Bruder, G.E., Kayser, J., Tenke, C.E., 2011. Event-related brain potentials in depression: clinical, cognitive, and neurophysiological implications. In: The Oxford Handbook of Event-Related Potential Components. Oxford University Press, p. 0. https://doi.org/10.1093/oxfordhb/9780195374148.013.0257.

Christ, S.E., Van Essen, D.C., Watson, J.M., Brubaker, L.E., McDermott, K.B., 2008. The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. Cereb. Cortex 19 (7), 1557–1566. https://doi.org/10.1093/cercor/bhn189.

Dahlke, J.A., Wiernik, B.M., 2019. Psychmeta: an R package for psychometric meta-analysis. Appl. Psychol. Meas. 43 (5), 415–416. https://doi.org/10.1177/0146621618795933.

Deng, X., Rosenfeld, J.P., Ward, A., Labkovsky, E., 2016. Superiority of visual (verbal) vs. auditory test presentation modality in a P300-based CIT: the Complex Trial Protocol for concealed autobiographical memory detection. Int. J. Psychophysiol. 105, 26–34. https://doi.org/10.1016/j.ijpsycho.2016.04.010.

Döring, N., Bortz, J., 2016. Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, *5. Aufl.* Springer.

Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R., Polich, J., Reinvang, I., Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clin. Neurophysiol. 120 (11), 1883–1908. https://doi.org/10.1016/j.clinph.2009.07.045.

Fazel, S., Danesh, J., 2002. Serious mental disorder in 23000 prisoners: a systematic review of 62 surveys. Lancet 359 (9306), 545–550. https://doi.org/10.1016/s0140-6736(02)07740-1.

Ford, J.M., Mathalon, D.H., 2004. Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. J. Psychiatr. Res. 38 (1), 37–46. https://doi.org/10.1016/s0022-3956(03)00095-5.

Funicelli, M., Salphati, S., Ungureanu, S., Laurence, J.R., 2023. Examining levels of processing using verbal & pictorial stimuli with the complex trial protocol in a mock theft scenario. Biol. Psychol. 183, 108666. https://doi.org/10.1016/j.biopsycho.2023.108666.

Garrigan, B., Adlam, A.L.R., Langdon, P.E., 2016. The neural correlates of moral decision-making: a systematic review and meta-analysis of moral evaluations and response decision judgements. Brain Cogn. 108, 88–97. https://doi.org/10.1016/j.bandc.2016.07.007.

Gati, I., Ben-Shakhar, G., 1990. Novelty and significance in orientation and habituation: a feature-matching approach. J. Exp. Psychol. Gen. 119 (3), 251–263. https://doi.org/10.1037/0096-3445.119.3.251.

Gigerenzer, G., 2004. Mindless statistics. J. Socio-Econ. 33 (5), 587–606. https://doi.org/10.1016/j.socec.2004.09.033.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley.

Grier, J.B., 1971. Nonparametric indexes for sensitivity and bias: computing formulas [doi:10.1037/h0031246]. Psychol. Bull. 75 (6), 424–429. https://doi.org/10.1037/h0031246.

Harrer, M., Cuijpers, P., Furukawa, T.A., Ebert, D.D., 2021. Doing Meta-analysis With R: A Hands-on Guide, 1st ed. Chapman & Hall/CRC Press https://www.routledge.com/Doing-Meta-Analysis-with-R-A-Hands-On-Guide/Harrer-Cuijpers-Furukawa-Ebert/p/book/9780367610074.

Hedges, L.V., Pigott, T.D., 2004. The Power of Statistical Tests for Moderators in Meta-analysis. American Psychological Association. https://doi.org/10.1037/1082-989X.9.4.426.

Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558. https://doi.org/10.1002/sim.1186.

Hu, X., Rosenfeld, J.P., 2012. Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. Psychophysiology 49 (8), 1090–1100. https://doi.org/10.1111/j.1469-8986.2012.01389.x.

Iacono, W.G., 2007. Detection of deception. In: Handbook of Psychophysiology, 3rd ed. Cambridge University Press, pp. 688–703. https://doi.org/10.1017/CBO9780511546396.029.

Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Stat. Med. 22 (17), 2693–2710. https://doi.org/10.1002/sim.1482.

Kok, A., 2001. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38 (3), 557–577. https://doi.org/10.1017/s0048577201990559.

Leue, A., Beauducel, A., 2019. A meta-analysis of the P3 amplitude in tasks requiring deception in legal and social contexts. Brain Cogn. 135, 103564. https://doi.org/10.1016/j.bandc.2019.05.002.

Lu, Y., Rosenfeld, J.P., Deng, X., Zhang, E., Zheng, H., Yan, G., Ouyang, D., Hayat, S.Z., 2018. Inferior detection of information from collaborative versus individual crimes based on a P300 Concealed Information Test. Psychophysiology 55 (4). https://doi.org/10.1111/psyp.13021.

Lüdecke, D., 2018. sjmisc: data and variable transformation functions. J. Open Source Softw. 3 (26), 754. https://doi.org/10.21105/joss.00754.

Lüdecke, D., 2023. sjPlot: Data Visualization for Statistics in Social Science. In R Package Version 2.8.17. https://CRAN.R-project.org/package=sjPlot.

Lukács, G., Weiss, B., Dalos, V.D., Kilencz, T., Tudja, S., Csifcsák, G., 2016. The first independent study on the complex trial protocol version of the P300-based concealed information test: corroboration of previous findings and highlights on vulnerabilities. Int. J. Psychophysiol. 110, 56–65. https://doi.org/10.1016/j.ijpsycho.2016.10.010.

Lykken, D.T., 1959. The GSR in the detection of guilt. J. Appl. Psychol. 43 (6), 385–388. https://doi.org/10.1037/h0046060.

Lykken, D.T., 1998. A tremor in the blood: uses and abuses of the lie detector. In: Plenum Trade. https://books.google.de/books?id=3sVdq5yqTUQC.

Matsuda, I., Nittono, H., Allen, J.J.B., 2013. Detection of concealed information by P3 and frontal EEG asymmetry. Neurosci. Lett. 537, 55–59. https://doi.org/10.1016/j.neulet.2013.01.029.

Meijer, E.H., Smulders, F.T., Johnston, J.E., Merckelbach, H.L., 2007. Combining skin conductance and forced choice in the detection of concealed information. Psychophysiology 44 (5), 814–822. https://doi.org/10.1111/j.1469-8986.2007.00543.x.

Meijer, E.H., Selle, N.K., Elber, L., Ben-Shakhar, G., 2014. Memory detection with the Concealed Information Test: a meta analysis of skin conductance, respiration, heart rate, and P300 data. Psychophysiology 51 (9), 879–904. https://doi.org/10.1111/psyp.12239.

Mertens, R., Allen, J.J.B., 2008. The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. Psychophysiology 45 (2), 286–298. https://doi.org/10.1111/j.1469-8986.2007.00615.x.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6 (7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Olson, J.M., Rosenfeld, J.P., Ward, A.C., Sitar, E.J., Gandhi, A., Hernandez, J., Fanesi, B., 2022. The effects of practicing a novel countermeasure on both the semantic and episodic memory-based complex trial protocols. Int. J. Psychophysiol. 173, 82–92. https://doi.org/10.1016/j.ijpsycho.2022.01.009.

Polich, J., 1989. Habituation of P300 from auditory stimuli. Psychobiology 17 (1), 19–28. https://doi.org/10.3758/BF03337813.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118 (10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Polich, J., Herbst, K.L., 2000. P300 as a clinical assay: rationale, evaluation, and findings. Int. J. Psychophysiol. 38 (1), 3–19. https://doi.org/10.1016/s0167-8760(00)00127-6.

Porjesz, B., Begleiter, H., 2003. Alcoholism and human electrophysiology. Alcohol Res. Health 27 (2), 153–160.

Rosenfeld, J.P., 2020. P300 in detecting concealed information and deception: a review. Psychophysiology 57 (7), e13362. https://doi.org/10.1111/psyp.13362.

Rosenfeld, J.P., Donchin, E., 2015. Resampling (bootstrapping) the mean: a definite do. Psychophysiology 52 (7), 969–972. https://doi.org/10.1111/psyp.12421.

Rosenfeld, J.P., Labkovsky, E., 2010. New P300-based protocol to detect concealed information: resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. Psychophysiology 47 (6), 1002–1010. https://doi.org/10.1111/j.1469-8986.2010.01024.x.

Rosenfeld, J.P., Cantwell, B., Nasman, V.T., Wojdac, V., Ivanov, S., Mazzeri, L., 1988. A modified, event-related potential-based guilty knowledge test. Int. J. Neurosci. 42 (1–2), 157–161. https://doi.org/10.3109/00207458808985770.

Rosenfeld, J.P., Soskins, M., Bosh, G., Ryan, A., 2004. Simple, effective countermeasures to P300-based tests of detection of concealed information. Psychophysiology 41 (2), 205–219. https://doi.org/10.1111/j.1469-8986.2004.00158.x.

Rosenfeld, J.P., Biroschak, J.R., Furedy, J.J., 2006. P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. Int. J. Psychophysiol. 60 (3), 251–259. https://doi.org/10.1016/j.ijpsycho.2005.06.002.

Rosenfeld, J.P., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., Chedid, E., 2008. The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. Psychophysiology 45 (6), 906–919. https://doi.org/10.1111/j.1469-8986.2008.00708.x.

Rosenfeld, J.P., Hu, X., Labkovsky, E., Meixner, J., Winograd, M.R., 2013. Review of recent studies and issues regarding the P300-based complex trial protocol for detection of concealed information. Int. J. Psychophysiol. 90 (2), 118–134. https://doi.org/10.1016/j.ijpsycho.2013.08.012.

Rosenfeld, J.P., Ozsan, I., Ward, A.C., 2017. P300 amplitude at Pz and N200/N300 latency at F3 differ between participants simulating suspect versus witness roles in a mock crime. Psychophysiology 54 (4), 640–648. https://doi.org/10.1111/psyp.12823.

Rosenfeld, J.P., Sitar, E., Wasserman, J., Ward, A., 2018. Moderate financial incentive does not appear to influence the P300 Concealed Information Test (CIT) effect in the Complex Trial Protocol (CTP) version of the CIT in a forensic scenario, while affecting P300 peak latencies and behavior. Int. J. Psychophysiol. 125, 42–49. https://doi.org/10.1016/j.ijpsycho.2018.02.006.

Rushby, J.A., Barry, R.J., 2009. Single-trial event-related potentials to significant stimuli. Int. J. Psychophysiol. 74 (2), 120–131. https://doi.org/10.1016/j.ijpsycho.2009.08.003.

Sai, L., Li, H., Wang, C., Rosenfeld, J.P., Lin, X., Fu, G., 2020. Feedback does not influence the recognition-related P300 in a novel concealed information test while feedback-evoked P300 shows promising diagnostic accuracy. Int. J. Psychophysiol. 157, 32–41. https://doi.org/10.1016/j.ijpsycho.2020.08.003.

Sasaki, M., Hira, S., Matsuda, T., 2001. Effects of a mental countermeasure on the physiological detection of deception using the event-related brain potentials. Shinrigaku Kenkyu 72 (4), 322–328. https://doi.org/10.4992/jjpsy.72.322.

Schauberger, P., Walker, A., 2025. openxlsx: Read, Write and Edit xlsx Files. In R Package Version 4.2.8. https://ycphs.github.io/openxlsx/index.html.

Schmidt, F.L., Hunter, J.E., 2015. Methods of Meta-analysis: Correcting Error and Bias in Research Findings, Third Edition ed. SAGE Publications, Ltd. https://doi.org/10.4135/9781483398105

Schooler, J., 2011. Unpublished results hide the decline effect. Nature 470 (7335), 437. https://doi.org/10.1038/470437a.

Siddle, D.A., 1991. Orienting, habituation, and resource allocation: an associative analysis. Psychophysiology 28 (3), 245–259. https://doi.org/10.1111/j.1469-8986.1991.tb02190.x.

Sokolov, E.N., 1963. Perception and the Conditioned Reflex. Macmillan.

Team, R. C, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Verschuere, B., Ben-Shakhar, G., 2011. Theory of the Concealed Information Test. In: Verschuere, B., Ben-Shakhar, G., Meijer, E. (Eds.), Memory Detection: Theory and Application of the Concealed Information Test. Cambridge University Press, pp. 128–148. https://doi.org/10.1017/CBO9780511975196.008.

Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. 36 (3), 1–48. https://doi.org/10.18637/jss.v036.i03.

Vrij, A., 2008. Detecting Lies and Deceit: Pitfalls and Opportunities, 2nd ed. John Wiley & Sons Ltd.

Ward, A.C., Rosenfeld, J.P., Sitar, E.J., Wasserman, J.D., 2020. The effect of retroactive memory interference on the P300-based Complex Trial Protocol (CTP). Int. J. Psychophysiol. 147, 213–223. https://doi.org/10.1016/j.ijpsycho.2019.10.016.

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. https://ggplot2.tidyverse.org.

Wickham, H., Bryan, J., 2025. readxl: Read Excel Files. In R Package Version 1.4.5. https://readxl.tidyverse.org.

Wickham, H., Hester, J., Bryan, J., 2024. readr: Read Rectangular Text Data. In R Package Version 2.1.5. https://readr.tidyverse.org.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D., 2025a. dplyr: A Grammar of Data Manipulation. In R Package Version 1.1.4. https://dplyr.tidyverse.org.

Wickham, H., Vaughan, D., Girlich, M., 2025b. tidyr: Tidy Messy Data. In R Package Version 1.3.1. https://tidyr.tidyverse.org.

Winograd, M.R., Rosenfeld, J.P., 2011. Mock crime application of the Complex Trial Protocol (CTP) P300-based concealed information test. Psychophysiology 48 (2), 155–161. https://doi.org/10.1111/j.1469-8986.2010.01054.x.

Winograd, M.R., Rosenfeld, J.P., 2014. The impact of prior knowledge from participant instructions in a mock crime P300 Concealed Information Test. Int. J. Psychophysiol. 94 (3), 473–481. https://doi.org/10.1016/j.ijpsycho.2014.08.002.

Zhang, J., Mueller, S.T., 2005. A note on ROC analysis and non-parametric estimate of sensitivity. Psychometrika 70 (1), 203–212. https://doi.org/10.1007/s11336-003-1119-8.