

Pre-Test Physiological Index Selection in the Concealed Information Test

Kohske Ogata¹

¹*Graduate School of Human Life & Ecology, Osaka Metropolitan University, Osaka, Japan*

Received 17.03.2026; Received revised 18.05.2026; Accepted 25.05.2026;

Available online 15.06.2026

The Concealed Information Test (CIT) is widely used in forensic contexts, particularly in Japan, to detect concealed knowledge through psychophysiological responses. Building on prior work on individual response specificity (IRS), this study examined whether pre-selecting diagnostic physiological indices based on performance in a preliminary demonstration (card) test influences the magnitude of subsequent CIT effects. Forty-six university students participated in a mock theft paradigm consisting of a card test and a watch trial. Skin conductance response (SCR), respiration speed (RS), and heart rate (HR) were recorded. Difference z-scores were used to index CIT effects. CIT effects were compared across three analytical patterns: using all three indices, using two indices excluding the weakest one, and using only the single most responsive index identified in the card test. Bayesian and frequentist repeated-measures analyses were conducted. Across both tasks, all three physiological indices showed positive CIT effects. Comparisons across index-selection conditions indicated differences in the magnitude of CIT effects. Results indicated that relying on a single, individually diagnostic index yielded numerically greater standardized within-subject differences between critical and noncritical items (i.e., larger CIT effects as indexed by difference z-scores) than combining multiple indices, although the associated Bayes factors and effect sizes were modest. Rather than focusing on detection accuracy, this study quantitatively examined differences in effect magnitude associated with index pre-selection using experimentally obtained within-participant data. The findings provide cautious support for IRS-informed index selection as a practical strategy for improving the interpretability of CIT outcomes.

Keywords: Concealed Information Test; individual response specificity; psychophysiology; index selection; Bayesian analysis.

Address of correspondence: Correspondence regarding this manuscript should be addressed to Kohske OGATA, Sugimoto 3-3-138, Osaka city, Japan. E-mail: ogata@omu.ac.jp

© 2026 Ogata, published by Sciendo

This work is licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Introduction

The Concealed Information Test (CIT) is a memory detection paradigm that relies on psychophysiological indices to infer the presence of crime-related knowledge (Iacono, 2024). In Japan, this psychological technique has been incorporated into daily forensic practice (Hira & Furumitsu, 2002). During a standard polygraph examination, the examinee first completes a preliminary procedure known as the demonstration test (or card test), after which the CIT concerning the crime-relevant details is administered (Matsuda et al., 2019).

Matsuda and colleagues (2006) proposed an approach to enhance the detection accuracy of the CIT by capitalizing on individual response specificity (IRS) in physiological measures. Specifically, they investigated how the weighting of different physiological indices, derived from the preliminary card test, could improve the discrimination between the presence and absence of memory traces. Using latent class analysis, they developed a classification procedure for memory detection that surpassed logistic regression and discriminant analysis in terms of accuracy. In that study, multiple indices were recorded, including heart rate (HR), systolic and diastolic blood pressure, cutaneous blood flow, respiratory time, respiratory amplitude, skin conductance level, and skin conductance response (SCR), and all were incorporated into the latent class model.

IRS refers to systematic interindividual differences in the dominant physiological channel through which concealed knowledge manifests during the CIT (Matsuda et al., 2015). Although the CIT is grounded in the orienting response framework and typically elicits enhanced electrodermal activity along with cardiac and respiratory changes (Meijer et al., 2014), prior research has demonstrated substantial variability in the relative strength of these indices across individuals (Gamer et al., 2008). For example, some examinees exhibit pronounced electrodermal differentiation, whereas others show more reliable modulation in heart rate or respiration (Gamer et al., 2008; Matsuda et al., 2015). Such variability provides a theoretical basis for considering whether physiological index selection might benefit from individualized calibration based on pre-test performance.

Subsequently, Matsuda et al. (2015) introduced a method to predict, on the basis of the demonstration test, whether an examinee would exhibit reliable physiological responses in the subsequent CIT. They focused on four measures—SCR, HR, respiration line length (RLL), and normalized pulse volume. Effect sizes were quantified using Cohen's *d*, and participants with effect sizes below 0.5 were excluded on the grounds of insufficient detectability. For the remaining cases, analyses revealed high area-under-the-curve values, indicating strong discriminability. Interestingly, although the combination of four indices

Pre-Test Physiological Index Selection in the Concealed Information Test

yielded the best overall accuracy, in some guilty participants the effect sizes for SCR or HR alone exceeded those of the composite index. In a related discussion, Matsuda et al. (2012) reviewed the historical reliance on aggregated z-scores across multiple physiological indices in CIT research, noting both the strengths and limitations of this approach. They emphasized that ignoring IRS and simply averaging across indices may obscure meaningful individual differences.

In Japanese forensic practice, the CIT has conventionally relied on three primary indicators: electrodermal activity, respiration, and cardiovascular responses (Hira & Furumitsu, 2002; Matsuda et al., 2019; Yamamura & Miyata, 1990). Typically, SCR serves as the electrodermal index, RLL or respiration speed (RS) as the respiratory measure, and HR as the cardiovascular measure (Meijer et al., 2014; Selle et al., 2017; Zaitzu, 2016). These three indices constitute the principal diagnostic measures. However, not all suspects display discernible responses across all three channels. Because IRS produces interindividual variability in reactivity, certain indices are more diagnostic for some individuals than for others.

Building on these insights, the present study sought to examine whether pre-selecting diagnostic physiological indices based on performance in the card test influences the magnitude of the within-subject CIT effect, as indexed by standardized difference. Specifically, I hypothesized that excluding indices with weak discriminability would amplify the canonical CIT pattern—characterized by enhanced SCR and reduced RS and HR (Meijer et al., 2014; Selle et al., 2017)—and thus increase the observed effect magnitude of the CIT as reflected in difference z-scores. In the context of deception detection, Studies 8 and 9 by Verschuere et al. (2023) demonstrated that detection performance was higher under a single cue condition than under a multiple cue condition, highlighting the advantage of focusing judgment on the most diagnostic cue. In contrast, within the framework of the CIT, Gamer et al. (2008) examined whether analyses should incorporate all three physiological indices—SCR, RS, and HR—or rely solely on SCR, which is often considered the most sensitive indicator. In Study 1, they constructed a logistic regression classifier, and in Studies 2 through 7, they evaluated its performance across six different datasets. Their results indicated that a composite function integrating all three indices yielded slightly higher detection accuracy than analyses based on a single index. Notably, however, the control group in Gamer et al.'s study was generated through simulation, and the outcomes were evaluated using receiver operating characteristic analyses rather than directly comparing standardized z-scores.

Importantly, ROC analyses evaluate classification performance across individuals (i.e., between-subject discrimination of guilty versus innocent examinees). In contrast, the present research question concerns within-subject signal structure—specifically, whether selecting physiological indices based on a preliminary task alters the magnitude of the differential response between critical and noncritical items within the same individual. Because IRS refers to variability in the dominant physiological channel through which concealed knowledge manifests, z-scores provide a more direct metric for examining amplification or attenuation of the CIT effect within persons. Thus, the present approach does not seek to replace ROC-based evaluation of diagnostic accuracy but rather to address a complementary, mechanistic question regarding signal magnitude.

The present study experimentally examines this issue. Importantly, the present study does not aim to evaluate diagnostic classification performance (e.g., accuracy, sensitivity/specificity, or ROC-based metrics). Instead, it tests whether prior index selection changes the magnitude of the CIT effect and quantifies these differences using standardized z-score indices. Accordingly, unlike Gamer et al. (2008), who generated a control group through simulation, the present study employed a within-subjects design in which all participants engaged in a mock theft paradigm, rather than a between-subjects design.

Methods

Participants

The study was conducted in accordance with procedures approved by the ethics committee of the author's affiliation (No.22-40). Undergraduate and graduate students were recruited and informed that participation would have no bearing on their academic evaluation and that they were free to withdraw at any time. Written informed consent was obtained from all participants prior to the experiment. A total of 46 students (34 women, 12 men), aged between 21 and 35 years, took part in the experiment.

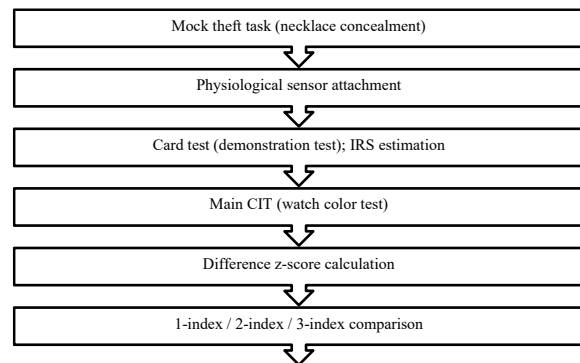


Figure 1. Overview of the experimental procedure and physiological index selection process.

Procedure

An overview of the experimental procedure and physiological index selection process is presented in Figure 1.

Mock theft task

After providing consent, participants were escorted to the experimental laboratory. They were instructed that the purpose of the task was to enact a mock theft. Specifically, after the experimenter moved to an adjacent room, participants were asked to open a white box placed before them, remove a necklace, and conceal it in one of the five drawers of a filing box. They were explicitly told not to open any of the drawers they had not chosen. Once the necklace was hidden, participants returned to the adjoining room where the experimenter was waiting.

Each drawer contained a wristwatch of a different color. Participants were not explicitly instructed to memorize the colors, but since the watches were placed at the front of each drawer, they inevitably saw them while hiding the necklace. In the subsequent CIT, the color of the concealed watch served as the critical item. This choice was based on prior evidence that the CIT can be successfully conducted using incidentally perceived stimuli rather than intentionally memorized information (Maoz et al., 2012;

Seymour & Fraynt, 2009), which is consistent with the focus of the present study.

Physiological recording setup

After this mock crime task, the polygraph sensors were attached and remained attached throughout the remainder of the experimental session.

Card test (demonstration test)

The session began with the card test (demonstration test), which was used as a pretest calibration procedure to estimate IRS across physiological indices. Participants were asked to select and memorize one card from a set of five; in reality, all cards were the “12 of hearts.” Thus, the critical item was “12 of hearts.” The buffer item was “joker,” and the noncritical items were “club 13,” “spade 11,” “heart 11,” and “spade 13.” The buffer item was always presented first in each block and was not included in the analyses. In CIT research and Japanese field-oriented procedures, initial buffer responses are commonly excluded to reduce nonspecific physiological responses associated with the beginning of a question sequence (e.g., Matsuda et al., 2012; Osugi, 2011; Zaitzu, 2016). During questioning, each item was visually presented on a display while the examiner asked aloud, and participants responded “no” to every item. The interval between questions was approximately 30 seconds. This sequence, consisting of six items (one critical, one buffer, four noncritical), was presented in randomized order and repeated across three blocks.

Main CIT (watch color test)

Following the card test, the watch trial was administered. Participants were questioned about the color of the watch that had been inside the drawer in which they had hidden the necklace. As in the card test, each color label was both displayed on screen and read aloud. The critical item was the actual watch color, the noncritical items were the other four colors, and “black” served as the buffer item. The set of six items was presented in randomized order and repeated three times. After completion of the CIT, the experimenter retrieved the necklace from the drawer, thereby confirming the actual color of the concealed watch.

Physiological Measures

Three classes of physiological indices were recorded. Cardiovascular activity was measured with an MT-11 system (NEC Medical Systems Inc., Japan) and amplified with an MT-31. HR was derived from the R-R interval, digitized at 1kHz. Electrodermal activity was measured with an AP-U030m device (DIGI-TEK Inc., Japan), and SCR was extracted for analysis. Respiration was measured using the Polyam-RESP system (Nihon Santeku Co., Ltd.).

SCR amplitude was defined as the difference between maximum and minimum conductance within 0.5–5.0s after stimulus onset and then log-transformed. RS was calculated as RLL during the 5.0–15.0s post-stimulus interval, divided by 10s. HR was computed as the average beats per minute during the 5.0–15.0s interval after stimulus presentation.

For each measure, z-scores were computed separately for critical and noncritical items within each block. Scores were then averaged across the three critical and 12 noncritical presentations. A difference z-score was derived by subtracting the mean noncritical z-score from the mean critical z-score. For comparability, RS and HR values were multiplied by -1 so that larger scores consistently reflected stronger CIT effects.

Data analysis

All statistical analyses were conducted using JASP (version 0.95.3). Difference z-scores served as the primary dependent variable, representing the standardized difference between responses to critical and noncritical items within each participant. For SCR, RS, and HR, z-scores were calculated separately within each block, averaged across the three blocks, and transformed so that higher values consistently indicated stronger CIT effects (RS and HR values were multiplied by -1). To examine whether physiological index selection influenced the magnitude of the CIT effect, three analytical conditions were compared within participants: the three-index condition, where difference z-scores were averaged across SCR, RS, and HR; the two-index condition, where difference z-scores were averaged across two indices after excluding the weakest index identified in the card test; and the single-index condition, where difference z-scores were derived from the index showing the highest difference z-score in the card test.

Because the same participants contributed data to all three analytical conditions, a repeated-measures analysis of variance (ANOVA) was conducted with the number of indices used for judgment as the within-subject factor and the difference z-score as the dependent variable. Both frequentist and Bayesian approaches were applied. For the Bayesian analysis, JASP default priors were used: a Cauchy distribution centered at zero with $r = 0.5$ for fixed effects, $r = 1.0$ for random effects, and $r = 0.354$ for covariates. Model priors assumed equal probability across competing models. Evidence for differences among conditions was evaluated using Bayes factors (BF). For the frequentist analysis, Mauchly's test was used to examine the assumption of sphericity, and Greenhouse–Geisser corrections were applied when necessary. Post hoc comparisons were performed using Holm-adjusted pairwise t-tests. Effect sizes were reported as Cohen's d for pairwise comparisons and partial η^2 (ηp^2) for omnibus effects.

Results

Figure 2 displays the difference z-scores obtained in the card test and the watch trial. Positive difference z-scores indicate the presence of a CIT effect, that is, stronger physiological responses to the critical item than to the noncritical items.

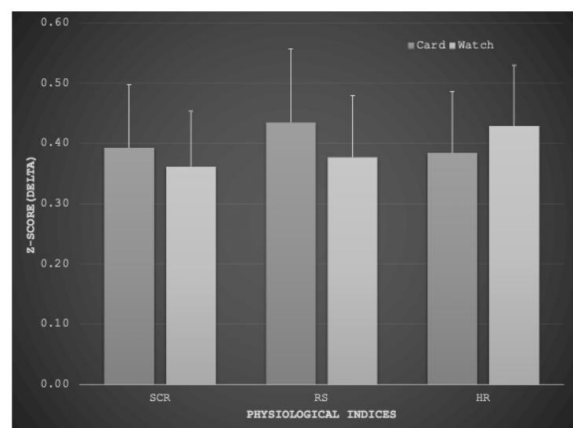


Figure 2. Difference z-scores for the card test and the watch trial across physiological indices.

Note. Bars represent mean difference z-scores (M) for SCR, RS, and HR in the card test and the watch trial. Error bars indicate standard errors.

Pre-Test Physiological Index Selection in the Concealed Information Test

Across both the card test and the watch trial, difference z-scores exceeded 0.4 for all three physiological indices—SCR, RS, and HR.

The present study further compared three analytical patterns in order to evaluate whether index selection enhanced the CIT effect: (a) scores derived from all three indices, (b) scores derived from the single index that showed the highest difference z-score in the card test, and (c) scores derived from two indices excluding the weakest index identified in the card test. For each participant, the physiological index showing the highest difference z-score in the card test was identified: 18 participants showed the strongest effect in SCR, 15 in RS, and 13 in HR. Figure 3 illustrates the results of this comparative analysis, and the specific combinations of selected indices are summarized in Table 1.

Table 1. Cross-tabulation of the most diagnostic indices identified in the card test (Pre-Test) and the watch trial (Post-Test).

		Post-Test		
		SCR	RS	HR
Pre-Test	SCR	8	3	7
	RS	5	6	4
	HR	2	3	8

Note. Pre-Test = card test; Post-Test = watch CIT. Values represent the number of participants whose most responsive index in the Pre-Test coincided with the indicated index in the Post-Test.

Inspection of Q-Q plots, skewness, kurtosis, and Shapiro–Wilk tests indicated no substantial deviations from normality across conditions (all p s > .12). A repeated-measures ANOVA was conducted to examine whether the magnitude of the difference z-scores varied across the three analytical patterns. The independent variable was the number of indices used for judgment, and the dependent variable was the difference z-score. For the effect sizes, the prior distribution followed the JASP default settings, namely a Cauchy distribution centered at zero. The r scale was set to 0.5 for fixed effects, 1.0 for random effects, and 0.354 for covariates. For the model prior, a uniform distribution was used, assuming equal prior probabilities for all models. The null hypothesis stated that the difference z-scores would be comparable regardless of the number of indices used for judgment, whereas the alternative

hypothesis assumed that the difference z-scores would vary as a function of the number of indices.

Bayesian analyses indicated that the model allowing differences among the three analytical patterns yielded a BF of 17.88, which, according to the criteria proposed by Kass and Raftery (1995), constitutes positive evidence in favor of differences among conditions. More specifically, post hoc comparisons revealed that the BF supporting a difference between the one-index and three-index conditions was 4.20. The BF comparing the one-index and two-index conditions was 7.24, whereas the BF for the comparison between the two-index and three-index conditions was 0.21. Frequentist analyses showed that Mauchly’s test indicated a violation of the sphericity assumption, and the Greenhouse–Geisser-corrected result was significant, $F(1.39, 62.71) = 6.87$, $p = .006$, $\eta^2 = .13$ (see Table 2). Holm-corrected pairwise comparisons revealed a significant difference between the one-index and three-index conditions, $t(45) = 2.73$, $p = .018$, Cohen’s $d = 0.37$, as well as between the one-index and two-index conditions, $t(45) = 2.96$, $p = .015$, $d = 0.32$ (see Table 3). In contrast, the comparison between the two-index and three-index conditions was not significant, $t(45) = 0.74$, $p = .465$, $d = 0.05$.

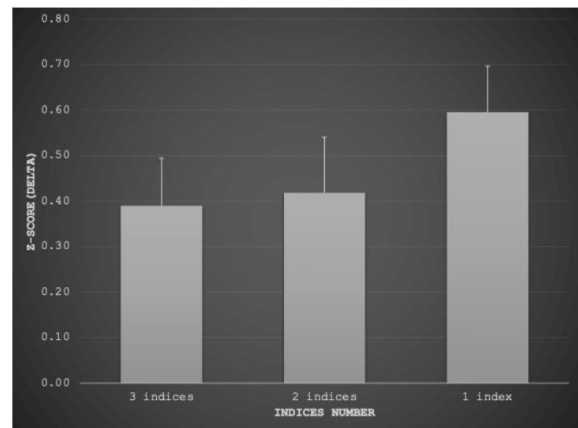


Figure 3. Difference z-scores in the watch trial as a function of the number of selected indices.

Note. Bars represent mean difference z-scores (Δ) for three analytical patterns: (a) all three indices, (b) two indices excluding the weakest one, and (c) the single most diagnostic index identified in the card test. Error bars indicate standard errors.

Table 2. Descriptive statistics and repeated-measures ANOVA results for difference z-scores across index selection conditions

Condition	M	SD
Three-index	.39	.42
Two-index	.42	.52
Single-index	.60	.69

Effect	F (df1, df2)	p	η^2
Number of indices used for judgment	6.87 (1.39, 62.71)	.006	.13

Note. M = mean difference z-score. SD = standard deviation. Greenhouse–Geisser correction was applied due to a violation of the sphericity assumption.

Table 3. *Pairwise comparisons of difference z-scores across index selection conditions*

Comparison			Mdiff	<i>t</i> (45)	<i>p</i>	Cohen's <i>d</i>	BF ₁₀
Single indices	vs.	Three	0.21	2.73	.018	0.37	4.20
Single indices	vs.	Two	0.18	2.96	.015	0.32	7.24
Two indices	vs.	Three	0.03	0.74	.465	0.05	0.21

Note. BF₁₀ represents the Bayes factor in favor of the alternative hypothesis relative to the null hypothesis. Positive values indicate larger difference z-scores in the first-listed condition.

Both Bayesian and frequentist analyses supported the presence of differences in difference z-scores across the three conditions. Post hoc comparisons further indicated that the mean standardized difference between critical and noncritical items (i.e., the within-subject CIT effect as indexed by difference z-scores) was numerically greater in the single-index condition than in the two- or three-index conditions. However, neither the Bayes factors nor the effect sizes were large, indicating that the magnitude of these differences was modest.

Discussion

The present study examined whether selecting physiological indices in advance, based on performance in the demonstration (card) test, could enhance the subsequent detection of concealed information. The findings suggest that excluding indices with weak discriminability was associated with differences in the magnitude of the CIT effect, as reflected in difference z-scores. In particular, analyses related to Figure 3 indicated that the standardized within-subject difference between critical and noncritical items (i.e., the CIT effect as indexed by difference z-scores) was numerically greater when only the most responsive physiological index for a given participant was used, compared with conditions in which multiple indices were combined. However, both the BF and the corresponding effect sizes indicated that these differences were modest in magnitude. These results are broadly consistent with earlier proposals emphasizing the role of IRS in psychophysiological detection paradigms (Matsuda et al., 2006, 2012, 2015). At the same time, the present study differs from prior IRS-oriented studies in that it focused not on classification accuracy or weighting procedures, but rather on whether pretest-based index selection could amplify the within-subject CIT effect itself. Thus, the present findings may provide preliminary evidence that physiological index pre-selection can influence the magnitude of concealed information effects even within the same examinee.

Although the present findings align with the IRS framework, it is important to note that the concordance between the most discriminative index in the card test and in the subsequent CIT was not perfect (Table 1). This pattern suggests that physiological dominance may not represent a fixed trait-like characteristic, but rather a

context-sensitive tendency that can vary depending on task demands and motivational factors (Gamer et al., 2008; Meijer et al., 2014). This interpretation is also consistent with the response fractionation perspective proposed by Klein Selle and Ben-Shakhar (2023), which suggests that different physiological indices in the CIT may reflect partially distinct cognitive and inhibitory processes rather than a single unitary detection mechanism. From this perspective, within-person variability in the “dominant physiological channel” across tasks may not necessarily indicate instability or measurement error, but rather differences in the relative contribution of orienting, attentional, and inhibitory processes across situations. Accordingly, while pre-test calibration may sharpen signal magnitude in some cases, it does not guarantee full cross-task stability of the dominant physiological channel. This limitation should be considered when interpreting the practical implications of index pre-selection.

Traditionally, CIT analyses have relied on averaged z-scores across multiple indices, without considering individual variability in responsiveness. While this approach offers a straightforward means of quantification, it risks diluting diagnostic information when certain indices are unresponsive in a given examinee. By contrast, the present results suggest that pre-selecting indices on the basis of the demonstration test may reduce such dilution and, in some cases, sharpen the signal of concealed knowledge, even though the overall magnitude of the effect was limited. Importantly, the present study focused specifically on the magnitude of within-subject physiological differentiation between critical and noncritical items rather than on categorical classification performance. Accordingly, the present findings should not be interpreted as suggesting that diagnostic accuracy is unimportant in applied CIT settings. Rather, the present study aimed to examine whether accounting for IRS could enhance the detectability of concealed information at the level of psychophysiological signal strength itself. The present findings should therefore be interpreted primarily within the context of CIT-based memory detection paradigms.

From a practical perspective, these findings may help to simplify CIT implementation in forensic contexts. In actual polygraph examinations conducted in Japan, three

indices—SCR, respiration, and HR—are typically employed (Hira & Furumitsu, 2002; Matsuda et al., 2019). Yet not all suspects produce reliable responses across all channels, which can complicate decision-making. At the same time, in Japanese field practice, the CIT is generally used not as a tool for directly classifying individuals as “guilty” or “innocent,” but rather as an investigative aid for estimating whether examinees recognize crime-related information (Matsuda et al., 2012; Nakayama, 2002; Osugi, 2011). In this sense, the present focus on the magnitude of differential physiological responding is broadly consistent with the practical logic of the CIT as a memory-detection procedure. The present results suggest that identifying, prior to the main test, the physiological index most likely to elicit a response in a given suspect may facilitate visual inspection-based judgments by reducing reliance on less informative indices, thereby offering potentially useful guidance for applied practice. Nevertheless, future studies should examine whether this type of index pre-selection also contributes to classification accuracy in paradigms where diagnostic decision-making is emphasized more directly.

Nevertheless, several limitations must be acknowledged. The relatively small sample size of the present study may have limited the strength of the conclusions that can be drawn regarding differences among index-selection conditions. In addition, the study employed a laboratory-based mock theft paradigm, which, although widely used, does not fully replicate the complexities of field investigations. Moreover, the current analyses focused on changes in the magnitude of the CIT effect rather than on diagnostic classification performance. Future research should therefore examine whether index pre-selection improves classification accuracy in applied forensic contexts. Further studies could also explore whether this approach generalizes to other physiological measures, such as blood pressure or neuroimaging-based indices, which have been proposed as potential markers in concealed information detection (e.g., Rosenfeld, 2005; Gamer, 2011).

Another important limitation concerns the absence of an innocent control group. The present study included only participants who engaged in a mock theft paradigm and therefore did not permit direct evaluation of false-positive rates or classification accuracy. Although this design is consistent with prior laboratory studies focusing on within-subject CIT effects and psychophysiological signal magnitude (e.g., Matsuda et al., 2006; Gamer et al., 2008), it limits conclusions regarding diagnostic performance. In particular, it remains possible that restricting analyses to a single index could increase the risk of misclassification in innocent examinees, especially if a non-diagnostic channel is selected. Future research should therefore incorporate innocent control groups to determine whether index pre-selection improves classification accuracy without elevating false-positive rates.

Conclusion

In conclusion, the present findings provide cautious support for the utility of pre-selecting diagnostic indices on the basis of a demonstration test to enhance the CIT effect. Although the observed effects were modest, this strategy may offer a potentially useful means of accounting for IRS and improving the interpretability of psychophysiological responses in both research and applied settings.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

The author has no relevant financial or non-financial interests to disclose. The author declares no conflicts of interest associated with this manuscript. No funding was received.

References

- Gamer, M. (2011). Detecting concealed information using autonomic measures. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and application of the Concealed Information Test* (pp. 27–45). Cambridge University Press. <https://doi.org/10.1017/CBO9780511975196.003>
- Gamer, M., Verschuere, B., Crombez, G., & Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiology & Behavior*, 95(3), 333–340. <https://doi.org/10.1016/j.physbeh.2008.06.011>
- Hira, S., & Furumitsu, I. (2002). Polygraphic examinations in Japan: Application of the guilty knowledge test in forensic investigations. *International Journal of Police Science & Management*, 4(1), 16–27. <https://doi.org/10.1177/146135570200400103>
- Iacono, W. (2024). Psychology and the lie detector industry: A fifty-year perspective. *Biological Psychology*, 190. <https://doi.org/10.1016/j.biopsycho.2024.108808>
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795. <https://doi.org/10.1080/01621459.1995.10476572>
- klein Selle, N., & Ben-Shakhar, G. (2023). A new theoretical perspective on concealed information detection. *Psychophysiology*, 60(3), e14187. <https://doi.org/10.1111/psyp.14187>
- Maoz, K., Breska, A., & Ben-Shakhar, G. (2012). Orienting response elicitation by personally significant information under subliminal stimulus presentation: Demonstration using the concealed information test. *Psychophysiology*, 49(12), 1610–1617. <https://doi.org/10.1111/j.1469-8986.2012.01470.x>
- Matsuda, I., Hirota, A., Ogawa, T., Takasawa, N., & Shigemasa, K. (2006). A new discrimination method for the concealed information test using pretest data and within-individual comparisons. *Biological Psychology*, 73(2), 157–164. <https://doi.org/10.1016/j.biopsycho.2006.01.013>
- Matsuda, I., Nittono, H., & Allen, J. J. (2012). The current and future status of the concealed information test for field use. *Frontiers in Psychology*, 3, 532. <https://doi.org/10.3389/fpsyg.2012.00532>
- Matsuda, I., Ogawa, T., Tsuneoka, M., & Verschuere, B. (2015). Using pretest data to screen low-reactivity individuals in the autonomic-based concealed

- information test. *Psychophysiology*, 52(3), 436–439. <https://doi.org/10.1111/psyp.12328>
- Matsuda I, Ogawa T and Tsuneoka M (2019) Broadening the Use of the Concealed Information Test in the Field. *Frontiers in Psychiatry*, 10, 24. <https://org.doi/10.3389/fpsy.2019.00024>
- Meijer, E. H., klein Selle, N., 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>
- Nakayama, M. (2002). Practical use of the concealed information test for criminal investigation in Japan. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 49–86). Academic Press.
- Osugi, A. (2011). Daily application of the CIT in Japan. In B. Verschuere, G. Ben-Shakhar, & E. Meijer (Eds.), *Memory detection: Theory and Application of the Concealed Information Test* (pp. 253–275). Cambridge University Press. <https://doi.org/10.1017/CBO9780511975196.015>
- Rosenfeld, J. P. (2005). ‘Brain fingerprinting’: A critical analysis. *Scientific Review of Mental Health Practice: Objective Investigations of Controversial and Unorthodox Claims in Clinical Psychology, Psychiatry, and Social Work*, 4(1), 20–37.
- Selle, N., Verschuere, B., Kindt, M., Meijer, E., & Ben-Shakhar, G. (2017). Unraveling the roles of orienting and inhibition in the concealed information test. *Psychophysiology*, 54(4), 628–639. <https://doi.org/10.1111/psyp.12825>
- Seymour, T. L., & Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Applied Psychophysiology and Biofeedback*, 34(3), 177–187. <https://doi.org/10.1007/s10484-009-9092-3>
- Verschuere, B., Lin, C. C., Huismann, S., Kleinberg, B., Willemse, M., Mei, E. C. J., van Goor, T., Löwy, L. H. S., Appiah, O. K., & Meijer, E. (2023). The use-the-best heuristic facilitates deception detection. *Nature Human Behaviour*, 7(5), 718–728. <https://doi.org/10.1038/s41562-023-01556-2>
- Yamamura, T., & Miyata, Y. (1990). Development of the polygraph technique in Japan for detection of deception. *Forensic Science International*, 44(2–3), 257–271. [https://doi.org/10.1016/0379-0738\(90\)90256-X](https://doi.org/10.1016/0379-0738(90)90256-X)
- Zaitsu, W. (2016). External validity of concealed information test experiment: Comparison of respiration, skin conductance, and heart rate between experimental and field card tests. *Psychophysiology*, 53(7):1100–1107. <https://doi.org/10.1111/psyp.12650>