Effects of Varying Numbers of Probes on RT-based CIT Accuracy

Jin-Sup Eom[1], Sunju Sohn[2], Kwangbai Park[1], Young-Ji Eum[3] and Jin-Hun Sohn[31]

[1]*Department of Psychology, Chungbuk National University, Cheongju, Korea*
[2]*Department of Welfare, Cheongju University, Cheongju, Korea*
[3]*Department of Psychology/Brain Research Institute, Chungnam National University, Daejeon, Korea*
[1]*jseom2003@hanmail.net,* [2]*ssohn124@cju.ac.kr,* [1]*kwangbai@chungbuk.ac.kr,*
[3]*petitaudrey@hanmail.net,* [3*]*jhsohn@cnu.ac.kr*

Abstract

RT-based Concealed Information Test (CIT) has been suggested to detect lying with high accuracy. However, because previous research have been conducted with a minimum of five probes, limited evidence is available to determine whether or not the RT-based CIT is also useful in precisely detecting lies with less than five probes. In this study, the accuracy of an RT-based Concealed Information Test (CIT) was examined by varying the numbers of probes used for the test. Results suggested that the RT-based CIT produces accurate lie detecting outcomes as the number of probes increases from a single probe to three probes. When five probes were used, however, the accuracy level did not improve from the level achieved with three probes. Interestingly, the accuracy decreased when the stimuli for the RT-based CIT were constructed with numeric elements of an event such as the amount of money. Further discussed are possible explanations in regards to the differences observed in the RT-based CIT accuracy rates dependent on the numbers of probes.

Keywords: Lie detection, Reaction time, Concealed information test, RT-based CIT

1. Introduction

Polygraph examination incorporates two major questioning techniques, one is the Comparison Question Technique (CQT); the other is the Concealed Information Test (CIT) [1]. CQT compares reactions to crime-specific questions (e.g., Did you steal money on x month/date at y location?) with those to other comparison questions (*e.g.*, Have you ever stolen money before 2012?). CIT, on the other hand, presents crime-relevant items of which the subject is assumed to be knowledgeable and several crime-irrelevant items. For example, if a knife was found in the murder scene, the knife and other weapons, including a gun, a rope, and an axe, are presented one at a time in a random order. When a subject's response to the crime-relevant item (*i.e.*, the knife) is systematically greater than that to the crime-irrelevant items, it is indicated that the subject identified the knife as the murder weapon used in the crime. This CIT is considered as an evidence-based strategy based on a theory of orienting response and habituation [2].

For CIT, physiological responses (*e.g.*, skin conductance response, blood pressure, and respiration, *etc.*,) are used to measure subjects' reaction to crime-relevant and irrelevant items. An alternative to this method is the use of electroencephalograms (EEG) in which event-related potential (ERP: brain

---

*Corresponding Author

potential linked with a certain stimulus) is measured 300 ms after exposure to each experimental stimulus [3-5]. Three types of stimuli have been most widely used in the P300-based CIT: one is the crime-relevant 'probe'; another is the 'irrelevant' that is not related to the crime; and the last is the 'target' intended to draw the subject's attention. Subjects are instructed to push 'yes' button immediately when the target appears and to push 'no' button when the probe or the irrelevant appears. To achieve a reliable P300, each stimulus type should be presented to the subject at least 20 times [6-7], and for P300-based CIT in general a minimum of 30 viewing sessions are repeated [8]. P300-based CIT is based on Johnson's (1986) triarchic model that assumes the amplitude of the P300 to become larger with significant and rare stimulus [9].

Recent studies have incorporated P300-based CIT paradigms with reaction time (RT) instead of EEG [10-14]. Suggested advantages of using an RT-based CIT are that only a single computer is required to administer the test, the test takes less than 10 minutes, and it yields relatively high lie detection accuracy. The RT-based CIT and P300-based CIT both share the same paradigm, with only difference being the constraint condition for the RT-based CIT in which subjects must voluntarily respond within one second after exposure to each stimulus.

RT-based CIT can be accounted for stimulus-response compatibility [15]. SRC is the degree to which what people perceive is consistent with the actions they need to take. SRC can be seen in the variation in the amount of time taken to respond to a visual stimulus, given the similarity of the event that prompts the action, and the action itself. For example, pushing the yellow button as a yellow light goes on produces higher SRC, compared to pushing a blue button when a yellow light goes on. Therefore, a high level of SRC is typically associated with a shorter reaction time, whereas a low level of SRC compatibility tends to result in a longer reaction time. In the case of criminals, probing with a crime-relevant object will result in a longer reaction time to produce a 'no' answer, compared to the time it takes to obtain a 'no' with an irrelevant object. This, in other words, suggests that if the response time to the relevant than the irrelevant is longer, the subject is likely to be lying.

According to Podlesny who reviewed the reports of FBI on polygraph investigations, probe items for CIT could be constructed for only 10.8% of the cases. Of all cases, 4.1% used only one or two probe items [16]. Adopting Lykken's suggestion of six or more probe items as the minimum requirement for administering CIT [17], only 2.1% of the entire cases were qualified for the test. Using a more generous criterion with three or more probes [18], the proportion of qualified cases increased to 6.7%. Based on Elaad's analysis of 98 CIT cases originally conducted by the Israel Police Scientific Interrogation Unit, 2.04 probe items on the average were used for each case [19].

The classical true score theory posits that each individual test score X is composed of a true score T and a measurement error score E, and test reliability is defined as the proportion of the variance in T to the variance in X. The more the test items, the smaller the variance in E; therefore, the test reliability generally increases with increasing number of test items [20]. A true score T is again composed of a construct score C and a systematic error score S. Test validity (*i.e.*, test accuracy) is operationalized as the proportion of the variance in C to the variance in X [21]. Therefore, the variance in E adversely impacts the test validity, which implies that test validity and test accuracy is proportional to the number of test items. This classical true score theory also applies to CIT.

In CIT examinations using physiological responses, the number of probes influences lie detection accuracy. In Ben-Shakhar and Elaad's study [22], one group was repeatedly presented twelve times with a single biographical item as the probe, while another group was shown four different biographical probes repeatedly for a total of three times, and the last group presented with twelve different biographical probes. Skin conductance

response (SCR) and respiration line length (RLL) data showed that lie detection accuracy improved as the number of probes increased. Ben-Shakhar and Elaad conducted a meta-analysis on 169 CIT testing data derived from 80 different laboratory studies [2]. Their findings suggested that a lie detection test is more accurate when at least five probe items are used, compared to a fewer number of probes.

Literature showing that the number of probes significantly affects lie detection accuracy of P300-based CIT is hard to come by. For example, Rosenfeld, Shue, and Singer used autobiographical information and mock crime information in their P300-based CIT study, where one group was tested three times on a single probe, while the other group was tested only once but with six different probes [23]. Their results are insufficient to suggest a notable difference in the ability to accurately discriminate subjects' guiltiness based on the number of probes. A review of ten P300-based CIT studies that were conducted based on the three types of stimulus paradigm (target, probe, and irrelevant), yet without incorporating a countermeasure, yielded inconsistent findings in regard to accuracy measured in association with the number of probes. Lie detection accuracy with a single probe ranged from 33% to 100%, while with six probes the accuracy ranged between 33% and 82%. The large variability in the reported accuracy may be due to the diversity of experimental environments.

Some P300-based CIT studies that measured response time as a means to maintain the subjects' attention to the stimuli showed differences in the response time between the probes and the irrelevant as a function of the number of probes [3-5, 13, 23-26]. Findings from these eight studies involving 15 experiments suggested that the difference in the response time between the probes and the irrelevant was significantly large when six probes were used (775.10 - 696.10 = 79.00 ms), compared to that of a single probe(570.37 - 536.64 = 33.73 ms). Of these 15 experiments, nine of them reported the mean and standard deviation of the response time, which were then used to calculate Cohen's d. From six studies that used a single probe, the Cohen's d was .54, compared to .86 calculated from three studies using six or more probes. An implication from these findings is that the total number of probes may also influence RT-based CIT results. However, RT-based CIT research, so far, have only tested the effects of five or six probes, which leaves a grey area as to the extent to which lie detection accuracy of the RT-based CIT is affected by the number of probes.

In the present study, an RT-based CIT was conducted by varying the number of probes from 1, 3, to 5 in Experiment 1 to see whether the lie detection accuracy differed respectively. In Experiment 2, the accuracy of the RT-based CIT was measured using two probes.

2. Experiment 1

2.1. Participants

Thirty-six college students (18 males) with normal or corrected normal vision were recruited as the participants by an advertisement promising 10,000 Korean won (equivalent to USD 9) for compensation. The mean age of the sample was 23.4 years (SD=2.6, range 19-28).

2.2. Apparatus

SuperLab (Version 4.0; Cedrus Corporation, San Pedro, CA) was used to present the stimuli on a 19" LCD monitor, and to measure and record response time.

2.3. Procedure

All participants initially performed a mock crime before being tested on a RT-based CIT. The participants were told that they would receive additional 10,000

won if they successfully beat the lie detection test following the mock crime. After the experiment, however, all participants were actually compensated with 20,000 won for participation regardless of their success.

The mock crime consisted of a scenario with five key elements: Enter the 'Psychophysiology lab' on the '5th floor' of the building, find a 'backpack', and run away with '150,000 Korean won' (equivalent to USD 132) and 'Kookmin credit card' from the 'backpack'. Participants were asked to remember the specific crime instructions given on a computer monitor. Once participants learned the mock crime instructions, they reviewed the instructions by answering questions on the five key elements (*i.e.*, 'psychophysiology lab', '5th floor', 'backpack', '150,000 won', and 'Kookmin credit card'): (1) which floor should you go to complete the crime?; (2) what is the name of the room to enter?; (3) what should you find in the room to steal money and credit card?; (4) how much money should you steal?; and (5) what is the name of the credit card you should run away with? The review test for the key elements of the mock crime instructions repeated three times before the participants were asked to actually commit the crime as instructed. Five minutes after the crime, the participants returned and acted as if they knew nothing about the crime experience.

Table 1. Test Materials

Crime Element	Probe	Target	Irrelevant
Floor	5th floor	2nd floor	1st floor 3rd floor 4th floor 6th floor
Lab. name	Psychophysiology	Neuroscience	Psychometrics Applied Psychology Behavior Observation Perception & Cognition
Object	Backpack	Desk	Jacket Handbag Table Chair
Money	150,000 won	300,000 won	50,000won 100,000won 200,000won 350,000won
Card	Kookmin card	Samsung card	Nonghyup Card Shinhan card Lotte card LG card

Note: All six stimuli for the lab. name were words with four syllables, two or three syllables for the object, and four syllables for the card.

2.4. RT-based CIT

Before conducting the RT-based CIT, the participants were asked to memorize the targets to respond affirmatively and answer memory questions about the targets three times. Questions included, 'What floor should you respond "yes" to?' 'What name of the room to enter to commit the crime will you respond "yes" to?' *etc*.

Participants were tested on the RT-based CIT three times repeatedly for the three experimental conditions. The order of the experimental conditions was counterbalanced among the participants with six participants for each order. One

condition used one of the five mock crime elements as a single probe (1P), another condition used three of the five mock crime elements (3P), and the other condition incorporated all five mock crime elements (5P). In each experimental condition, stimuli in the same element category were used as the probe, the target, and the irrelevants (see Table 1).

In all conditions, the relative frequencies with which the three types stimulus were presented to the participant were fixed to be 1:1:4 for the probe, the target, and the irrelevant respectively. Stimuli in 1P consisted of one probe, one target, and four irrelevants, and they were presented to the participants in 30 blocks with 6 trials (*i.e.*, 6 sequences of stimulus-response) per block. Stimuli in 3P were composed of three probes, three targets, and twelve irrelevants that were presented in 10 blocks with 18 trials per block. Lastly, in 5P, five probes, five targets, twenty irrelevants were presented in 6 blocks each of which consisted of 30 trials. In each block of trials, the probes, the targets, and the irrelevants were presented in a random order. When the target appears, the participants pressed the 'yes' button (left mouse button). They pressed 'no' (right mouse button) for the rest of the stimuli. Following approximately 90 trials, participants were provided 30 seconds to rest.

A stimulus, written in white Korean letters of 2 cm height on black background, appeared at the center of the monitor screen up to a maximum of 1,000 ms and disappeared as soon as the participant responded. If the participant did not respond within 1,000 ms, 'too slow' statement appeared on the screen for one second. Intervals between two consecutive trials were 800, 900, 1,000, 1,100, or 1,200 ms.

2.5. Results

Reaction time was examined using two-way repeated measure ANOVA, with the number of probes and stimulus type as independent variables. The main effects of both the number of probes, $F(2,70)=110.004$, $p<.001$, and stimulus type, $F(1,35)=49.895$, $p<.001$, and the effect of interaction between the two variables, $F(2,70)=21.792$, $p<.001$, were all significant. Findings from simple main effect analyses indicated significant differences in the reaction time between the two stimulus types within the conditions of three probes (3P), $F(1,35)=44.903$, $p<.001$, and five probes (5P), $F(1,35)=34.965$, $p<.001$. As shown in Figure 1, reaction times were longer for probe than irrelevant in 3P and 5P. The reaction time did not differ between the two stimulus types within the condition of a single probe (1P), $F(1,35)=.016$, $p>.5$.

Differences in the reaction time by stimulus type (probe versus irrelevant) were further explored by conducting paired t-test for each element of the mock crime. As Table 2 shows, with a single probe (1P), a significant difference between the two types of stimulus was found only for the 'floor' element, $t(7)=-2.575$, $p<.05$, where the response time following the irrelevant stimuli were longer than that following the probes. With three probes (3P), differences between the two types of stimulus were detected for the name of the lab, $t(23)=4.688$, $p<.001$, the specific object ('backpack'), $t(23)=5.743$, $p<.001$, and the name of the card, $t(17)=2.596$, $p<.05$. Five probes (5P) also yielded comparable results: differences in reaction time between the two types of stimuli were present for the name of the lab, $t(35)=6.723$, $p<.001$, the specific object, $t(35)=6.487$, $p<.001$, and the name of the card, $t(35)=3.287$, $p<.01$.
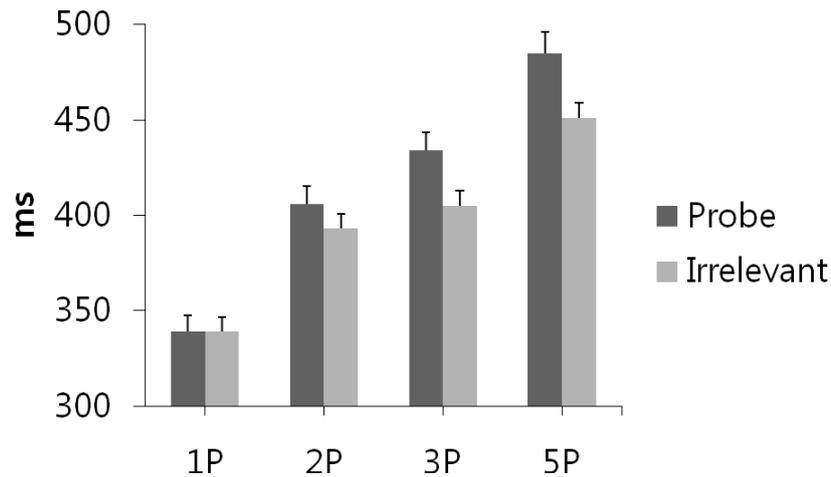
Figure 1. Mean Reaction Times for each Experimental Condition. Error Bars Represent Standard Error of the Mean

Table 2. Mean Reaction Times (in ms) for each Crime Element

Condition	Floor		Lab. name		Object		Money		Card	
	Probe	Irrelevant	Probe	Irrelevant	Probe	Irrelevant	Probe	Irrelevant	Probe	Irrelevant
1P	321* (28)	330 (24)	372 (69)	363 (66)	356 (45)	359 (49)	322 (59)	327 (58)	335 (28)	334 (25)
3P	401 (45)	409 (46)	465* (88)	414 (61)	452* (57)	399 (47)	432 (67)	414 (59)	431* (76)	405 (47)
5P	462 (78)	457 (54)	542* (69)	474 (58)	521* (106)	439 (52)	456 (71)	451 (45)	481* (90)	446 (45)
2P	381* (42)	395 (48)	454 (40)	437 (38)	404* (66)	369 (45)	376* (28)	390 (35)	433* (70)	395 (38)

Note: Standard deviations are given between parentheses
* $p<.05$

2.6. Discussion

In Experiment 1, differences in reaction time between two types of stimuli differed as a function of the number of probes. With a single probe, there was no significant difference in the response time between the probe and irrelevant stimuli. However, the difference emerged with three and five probes. The effect sizes of stimulus type were comparable in the three probes condition (d=1.16) and five probes condition (d=1.19). With three and five probes, the differences in reaction time were found for the stimuli that involved names of a location (*i.e.*, Psychophisiology Lab), a specific object (*i.e.*, backpack), and a credit card (Kookmin), but not for the stimuli containing numbers (i.e., floor and amount of money). These findings may suggest that stimulus-response incompatibility is associated only with word-specific contents.

3. Experiment 2

Based on the Experiment 1, in the second experiment we measured the accuracy of the RT-based CIT was measured using two probes.

3.1. Participants

The sample included 30 college students (15 males) with normal or corrected normal vision. The mean age of the sample was 23.5 years (SD=1.9, range 20-27).

3.2. Procedure

The same mock crime, instructions, procedure, and apparatus as used in Experiment 1 were also used in this experiment.

3.3. RT-based CIT

Prior to testing, as did the participants in Experiment 1, the participants were asked to memorize the targets to respond affirmatively and answer memory questions about the targets three times.

Subjects were administered the RT-based CIT using two (2P) of the five probes. Stimuli in 2P consisted of two probes, two targets, and eight irrelevants, and they were presented to the participants in 15 blocks with 12 trials per block. In each block of 2P, each of the twelve stimuli was presented once. Testing and stimulus presentation methods were identical with those in Experiment 1.

3.4. Results

Reaction time was analyzed using one-way repeated measure ANOVA, with stimulus type as an independent variable. The effects of stimulus type, $F(1,29)$=5.017, p<.05, were significant. As shown in Figure 1, reaction time was longer for probe than irrelevant in 2P.

Differences in the reaction time between stimulus types (probe versus irrelevant) were further explored by paired t-test for each of the five key elements of the mock crime (Table 2). Significant differences were found for the floor, $t(11)$=-2.412, p<.05; the specific object ('backpack'), $t(11)$=3.775, p<.01; the amount of money, $t(11)$=-2.705, p<.05; and the name of the credit card, $t(11)$=2.515, p<.05. This finding on the reaction time for the floor element and the amount of money was contrary to our expectations.

3.5. Discussion

Experiment 2 was to see whether or not RT-based CIT with two probes could be interpolated into the findings from Experiment 1. The difference in the response time between the probe and the irrelevant with two probes was 13 ms with an effect size d=0.42. This finding indicates that as the number of probes increases from 1, 2, to 3, the difference in the response time between the probe and the irrelevant, i.e., the size of the effect of stimulus type on reaction time, may also increase.

4. General Discussion

This study examined whether or not the number of probes impacts lie detection accuracy of RT-based CIT. Results showed that the effect sizes of stimulus type (probe versus irrelevant) were close to zero (Cohen's d = -.02) with a single probe, d=0.42 with two probes, d=1.16 with three probes, and d=1.19 with five probes. Overall, the sensitivity of the test systematically improved as the number of probes

increased from one to three, although there was no notable difference between three and five probes. Our finding contradicts Seymour and Kerlin who suggested that the number of probes is not likely to influence test efficiency [11].

Ben-Shakhar and Elaad reviewed 169 CIT experiments and showed that the effect size was $d=1.29$ with four or less CIT probes, but it increased to $d=2.35$ with five or more probes [2]. More specifically, the mean effect size was $d=1.29$ with a single probe, $d=1.34$ with two probes, $d=.96$ with three probes, and $d=1.19$ with four probes, indicating that the accuracy of CIT would not meaningfully change with different numbers of probes less than or equal to four. In this study, however, we observed differences in effect sizes with the number of probes less than or equal to three. This inconsistency between the findings from the previous and the present studies is probably due to a difference in the paradigm adopted in CIT and RT-based CIT. CIT relies on orienting response and habituation (Ben-Shakhar & Elaad, 2002a) [27], while RT-based CIT is based on stimulus-response compatibility [15].

According to Verschuere and Houwer [15], the behavior of innocent participants in RT-based CIT depends mostly on stimulus familiarity: They respond differently to familiar targets and unfamiliar probes and irrelevants. Guilty participants, on the other hand, distinguish targets and probes from irrelevants because of their differential familiarity. The validity of the test depends on the premise that guilty participants have to response as if they are unfamiliar or unaware of the probes, which is likely to result in stimulus-response incompatibility. Since guilty participants should produce a response required for unfamiliar stimuli to familiar probes, they will take relatively longer time to process the probes compared to other irrelevant stimuli. Given this logic, the lack of significant difference in the response time between the probe and the irrelevant when one or two probes were used in the present study can be interpreted that stimulus-response incompatibility is least detectable under these conditions.

The effects of the number of stimulus-set are also reported for Stroop color word test and picture-word interference task [28]. In Williams' study, Stroop test was conducted by measuring the response time with increasing numbers of colors and color names from two to eight [29]. Results showed that the more diverse the colors, the longer it takes to accurately identify colored words. Williams suggested that this is because people are more likely to experience perceptual conflict with more diversity in colors. Similarly, La Heji and van den Hof showed that the time taken by subjects to name each picture was closely related to the size of a target-set in picture-word interference [28]. That is, there is less picture-word interference with a smaller size of target-set because it is easier for subjects to retrieve target-names with fewer distractions. Findings from the present study can also be understood as resulted from perceptual conflict or difficulty in memory retrieval.

An interesting aspect of our findings is that the elements of mock crime in the experiment (*i.e.*, floor, names of place and card, specific object, and amount of money) yielded diverse differences in response-time. For example, the difference in the response time between the probe and the irrelevant with respect to the place name, the specific object, and the card name in the 3P and 5P conditions were significant. Differences were also observed in the 2P condition with the place name and the card name. However, no significant difference was observed with the floor or the amount of money in any of the conditions. In addition, the response time in 1P and 2P conditions was in fact significantly shorter for the probes than for the irrelevants with respect to the floor and the amount of money. The key difference among these elements is that the floor and the amount of money were numeric stimuli while the names of the place, the object, and the card were non-

numeric words. A similar phenomenon was also observed in Gamer, Bauermann, Stoeter, and Vossel [30], where there was a significant difference in the response time between the probe and the irrelevant when playing cards (probe and target: jack and king of spaces respectively) were the stimuli. Bank notes (probe and target: 20 and 40 EUR bank note respectively), however, did not produce noteworthy differences in the response time. Gamer *et al.*, suggested that the phenomenon, *i.e.*, the lengthened response time to jack of spaces, was due to a larger perceptual similarity between the playing cards that increased response conflict [30]. Results from the present study could be understood from a similar point of view. That is, compared to numbers, Korean words are consisted of cognitively complex consonants and vowels. However, it is yet to be tested whether cognitive complexity raises the degree of response conflict as the perceptual similarity does.

Acknowledgments

References

[1] D. T. Lykken, "The GSR in the detection of guilt", Journal of Applied Psychology, vol. 43, no. 6, **(1959)**, pp. 358-388.

[2] G. B. Shakha and E. Elaad, "The validity of psychophysiological detection of information with the guilty knowledge test: A meta-analytic review", Journal of Applied Psychology, vol. 88, no. 1, **(2003)**, pp. 131-151.

[3] L. A. Farwell and E. Donchin, "The truth will out: Interrogative polygraphy ('lie detection') with event-related brain potentials", Psychophysiology, vol. 28, no. 5, **(1991)**, pp. 531-547.

[4] J. J. Allen, W. G. Iacono and K. D. Danielson, "The identification of concealed memories using the event-related potential and implicit behavioral measures: A methodology for prediction in the face of individual differences", Psychophysiology, vol. 29, no. 5, **(1992)**, pp. 504-522.

[5] R. Mertens and J. J. B. Allen, "The role of psychophysiology in forensic assessments: Deception detection, ERPs, and virtual reality mock crime scenarios", Psychophysiology, vol. 45, no. 2, **(2008)**, pp. 286-298.

[6] J. Polich, "P300 development from auditory stimulus", Psychophysiology, vol. 23, no. 5, **(1986)**, pp. 590-597.

[7] J. Cohen and J. Polich, "On the number of trials needed for P300", International Journal of Psychophysiology, vol. 25, no. 3, **(1997)**, pp. 249-255.

[8] J. P. Rosenfeld, M. Soskins, G. Bosh and A. Ryan, "Simple, effective countermeasures to P300-based tests of detection of concealed information", Psychophysiology, vol. 41, no. 2, **(2004)**, pp. 205-219.

[9] R. Johnson, "A triarchic model of P300 amplitude", Psychophysiology, vol. 23, no. 4, **(1986)**, pp. 367–384.

[10] T. L. Seymour, C. M. Seifert, M. G. Shafto and A. L. Mosmann, "Using response time measures to assess "Guilty Knowledge", Journal of Applied Psychology, vol. 85, no. 1, **(2000)**, pp. 30-37.

[11] T. L. Seymour and B. R. Fraynt, "Time and encoding effects in the concealed knowledge test", Applied Psychophysiology and Biofeedback, vol. 34, no. 3, **(2009)**, pp. 177-187.

[12] T. L. Seymour and J. R. Kerlin, "Successful detection of verbal and visual concealed knowledge using an RT-based paradigm", Applied Cognitive Psychology, vol. 22, no. 4, **(2008)**, pp. 475-490.

[13] B. Verschuere, G. Crombez, T. Degrootte and Y. Rosseel, "Detecting concealed information with reaction time: Validity and comparison with the polygraph", Applied Cognitive Psychology, vol. 24, no. 7, **(2010)**, pp. 991-1002.

[14] G. V. Petra, M. Miclea and L. V. Petra, "Reaction time-based detection of concealed information in relation to individual differences in executive functioning", Applied Cognitive Psychology, vol. 26, no. 3, **(2012)**, pp. 342-351.

[15] B. Verschuere and J. D. Houwer, "Detecting concealed information in less than a second: response latency-based measures", Edited B. Verschuere, G. Ben-Shakhar, and E. Meijer, Memory detection: Theory and Application of the Concealed Information Test, Cambridge University Press, London, **(2011)**, pp. 46-62.

[16] J. A. Podlesny, "A paucity of operable case facts restricts applicability of the guilty knowledge technique in FBI criminal polygraph examinations", Forensic Science Communications, Retrieved from http://www2.fbi.gov/hq/lab/fsc/backissu/july2003/podlesny.htm, vol. 5, no. 3, **(2012)**.

[17] D. T. Lykken, "Detection of guilty knowledge: A comment on Forman and McCauley", Journal of Applied Psychology, vol. 73, no. 2, **(1988)**, pp. 303-304.

[18] R. F. Forman and C. McCauley, "Validity of the positive control polygraph test using the field practice model", Journal of Applied Psychology, vol. 71, no. 4, **(1986)**, pp. 691-698.

[19] E. Elaad, "Detection of guilty knowledge in real-life criminal investigations", Journal of Applied Psychology, vol. 75, no. 5, **(1990)**, pp. 521-529.

[20] L. Croker and J. Algina, "Introduction to classical and modern test theory", Harcourt Brace Jovanovich, New York, **(1986)**.

[21] D. L. Streiner, "Starting at the beginning: An introduction to coefficient alpha and internal consistency", Journal of Personality Assessment, vol. 80, no. 1, **(2003)**, pp. 99-103.

[22] G. B. Shakhar and E. Elaad, "Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: A reexamination", Journal of Applied Psychology, vol. 87, no. 5, **(2002)**, pp. 972-977.

[23] J. P. Rosenfeld, E. Shue and E. Singer, "Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information", Biological Psychology, vol. 74, no. 3, **(2007)**, pp. 396-404.

[24] M. Gamer and S. Berti, "Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials", Psychophysiology, vol. 47, no. 2, **(2010)**, pp. 355-364.

[25] E. H., Meijer, F. T. Y. Smulders and A. Wolf, "The contribution of mere recognition to the P300 effect in a concealed information test", Applied Psychological Biofeedback, vol. 34, no. 3, **(2009)**, pp. 332-226.

[26] J. P. Rosenfeld, J. R. Biroschak and J. J. Furedy, "P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms", International Journal of Psychophysiology, vol. 60, no. 3, **(2006)**, pp. 251-259.

[27] G. B. Shakhar and E. Elaad, "The guilty knowledge test (GKT) as an application of psychophysiology: Future prospects and obstacles", Edited M. Kleiner, Handbook of Polygraph Testing, Academic Press, San Diego, **(2002)**, pp. 87-102.

[28] W. L. Heij and E. V. D. Hof, "Picture-word interference increases with target-set size", Psychological Research, vol. 58, no. 2, **(1995)**, pp. 119-133.

[29] E. Williams, "The effects of amount of information in the Stroop Color word test", Perception & Psychophysics, vol. 22, no. 5, **(1977)**, pp. 463-470.

[30] M. Gamer, T. Bauermann, P. Stoeter and G. Vossel, "Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information", Human Brain Mapping, vol. 28, no. 12, **(2007)**, pp. 1287-1301.