

## Effects of Consistency Criterion for Scoring on the Reliability and the Validity of Polygraph Test for Crime Suspects\*

### 범죄 용의자의 거짓말탐지검사의 신뢰도와 타당도에 대한 일관성 채점기준의 효과

Yu-hwa Han\*\* · Je-young Jeong\*\*\* · Kwangbai Park\*\*†

한유화\*\* · 정재영\*\*\* · 박광배\*\*†

Department of Psychology, Chungbuk National University\*\*

충북대학교 심리학과\*\*

Prosecutors' Office of the Republic of Korea\*\*\*

대한민국 대검찰청 심리분석실\*\*\*

#### Abstract

For scoring polygraph charts, the Prosecutors' Office of the Republic of Korea uses a consistency criterion in which an elevated signal on one physiological channel is scored as a deceptive response only if the signal is also elevated on other channels. In the current study, the effects of this scoring criterion on reliability and accuracy (validity) of polygraph scores were assessed. Polygraph tests on 26 suspects were evaluated twice by the same examiners. The examiners used the consistency criterion in the first evaluation. In the second evaluation, the examiners were prevented from using the criterion; the signals from each physiological channel were separated and randomly arranged before they were rescored by the same examiner. Reliability was assessed by the variation among the scores for each suspect. Accuracy was assessed by establishing a standard, based on a Latent Class Analysis model, using the results of polygraph tests on each of 182 additional suspects. Reliability and accuracy were both improved by the use of the consistency criterion which therefore was recommended.

**Keywords :** Polygraph, Comparison Question Test, Scoring criterion, Consistency

#### 요약

현재 한국에서는 거짓말탐지검사의 채점과정에서 '일관성' 기준을 사용하고 있다. 본 연구에서는 현재 한국에서 사용하고 있는 거짓말탐지검사 기법의 채점 기준 중 '일관성' 기준의 실증적 타당성을 검증하고자 하였다. 구체적으로는 한국에서 이루어지고 있는 거짓말탐지검사의 채점과정에서 '일관성'을 실제로 고려하고 있는지의 여부를 검증한 후 '일관성'을 고려하여 채점하는 경우와 그렇지 않은 경우에서 거짓말탐지검사 결과의 신뢰도(Cronbach'  $\alpha$ )와 정확성이 어떻게 달라지는지를 파악하였다. 본 연구는 실제 현장에서 '일관성'을 고려하여 채점 한 원채점자료와 '일관성'을 고려할 수 없도록 가공하여 채점한 실험채점자료를 분석하였다. 원채점자료와 실험채점자료는 동일한 검사관들이 채점하도록 하였다. 그 결과 현재 한국에서는 거짓말탐지검사의

\* 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(M10740030003-07N4003-00310).

† 교신저자 : 박광배 (충북대학교 사회과학대학 심리학과)

E-mail : kwangbai@chungbuk.ac.kr.

TEL : 043-261-2195

채점과정에서 ‘일관성’을 고려하고 있는 것이 확인되었다. 신뢰도는 원채점자료가 .93, 실험채점자료가 .91로 비슷한 수준이었다. 정확성은 검찰의 기소 여부와 거짓말탐지검사 점수에 의한 판정을 지표로 사용한 잠재계층분석(N=182)의 결과를 실제 거짓말 여부를 판단하는 기준으로 사용했을 때 원채점자료에서 76.9%, 실험채점자료에서 61.5%로 나타났다. 논의에서는 이러한 결과가 나타난 원인과 한국에서 사용하고 있는 ‘일관성’ 기준이 거짓말탐지검사 채점에서 갖는 유용성에 대하여 논의하였다.

**주제어 :** 거짓말탐지기, 비교질문검사, 채점기준, 일관성

## 1. Introduction

A polygraph test for lie detection involves a scoring procedure which is undertaken by trained experts. Although the scoring procedure is standardized with a detailed set of criteria, the application of the criteria relies rather heavily on subjective judgments of the scorer.<sup>1)</sup> In addition to the standard set of scoring criteria, which was developed by the Department of Defense Polygraph Institute (DoDPI) of the United States, the Prosecutors' Office of the Republic of Korea uses a "consistency criterion" (CC) for scoring polygraph charts obtained with Comparison Question Test (CQT).<sup>2)</sup> According to the CC, an elevated signal from one physiological channel is scored as a lie response only if the signals from other channels are also elevated.

The CC is specifically applied in the following fashion. Each examinee is tested three times repeatedly with an identical set of two relevant and three comparison questions, yielding three charts of signals emanating from four physiological channels. The signals corresponding to each of the relevant questions are the spots to be scored, resulting in 12 scoring spots for each relevant question. Suppose that the signal at a scoring spot corresponding to the first relevant question is consistent with the DoDPI criteria for a lie response. This spot can be scored as a lie response only if at least 8 of the remaining 11 scoring spots corresponding to the same relevant question are also consistent with the

DoDPI criteria for lie responses.

The CC is considered to help reduce the false positive decision errors. In addition to the telling of a lie, signals from physiological channels can be elevated by a host of random factors. If a signal from one physiological channel is elevated by a lie then similarly elevated signals may emanate from other channels if the polygraph test is internally consistent. Therefore, the CC is believed to help prevent misinterpretations that are caused by extraneous factors. This rationale for the CC is supported if two conditions are met: (1) telling a lie elevates signals from all physiological channels simultaneously and consistently; (2) extraneous factors such as normal test anxiety, body movement, and sensor slippage affect physiological signals in a random fashion elevating some signals sometimes. If the former condition is not met, the application of the CC may inflate the rate of false negatives, erroneously judging a deceptive examinee to be truthful. If the latter condition is not met, it may inflate the rate of false positives, erroneously judging a truthful examinee to be deceptive. Whether the two conditions are truly met in the application of polygraph tests to criminal investigations is unknown.

The CC can also affect the reliability of the polygraph scores. In practice, the CQT is usually conducted at least three times with an identical set of questions. The CC can enhance the test-retest reliability of the polygraph test by making the scores on the successive tests more dependent on one another. On the other hand, it is well known that physiological responses tend to be reduced when an organism is stimulated repeatedly. If the examinee tells a lie consistently, the strength of the signals may become weaker or less stable on later test due to physiological adaptation. Increased instability of

1) Iacono, W. G. & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinions. *Journal of Applied Psychology*, 82, 426-433.

2) Raskin, D. C. & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 1-47). San Diego: Academic Press.

Table 1. Descriptive Statistics of the Score Sets

Score Set	Indictment	Original Polygraph Decision	n	Subtotal (sum of 8 spot scores)			Grand Total
				First Test Mean (SD)	Second Test Mean (SD)	Third Test Mean (SD)	
Original	Indicted	DI*	8	-6.63 (2.13)	-6.88 (2.42)	-6.88 (2.36)	-19.38 (6.09)
		NDI**	1	1	-1	0	0
		Inconclusive	0	—	—	—	—
	Dropped	DI*	8	-5.38 (1.19)	-5.13 (2.03)	-4.50 (1.51)	-15.00 (1.85)
		NDI**	4	-0.25 (1.71)	1.25 (1.50)	2.50 (1.29)	3.50 (2.89)
		Inconclusive	5	-0.40 (0.89)	-0.40 (0.89)	-1.40 (1.34)	-2.20 (0.45)
Segmented	Indicted	DI*	8	-6.38 (2.26)	-4.88 (2.7)	-3.50 (4.26)	-14.75 (7.73)
		NDI**	1	4	2	0	6
		Inconclusive	0	—	—	—	—
	Dropped	DI*	8	-4.94 (1.99)	-4.5 (3.05)	-3.88 (2.18)	-13.31 (5.70)
		NDI**	4	1.38 (3.28)	2.00 (2.48)	2.63 (2.66)	5.88 (6.09)
		Inconclusive	5	-0.40 (2.99)	0.40 (1.19)	-2.40 (2.82)	-2.40 (4.20)

\*DI: deception identified; \*\*NDI: n deception identified.

signals due to physiological adaptation may cause an inconsistency in scoring on successive tests especially when the CC is applied.

The present study examined the consequences of the application of the CC on the reliability and the accuracy (validity) of scores on the CQT for crime suspects.

## 2. Method

### 2.1. Cases

Twenty-six actual cases that were tested in previous years were retrieved from archives of the Prosecutors' Office of the Republic of Korea. These cases were selected for the present study because they could be rescored by the original examiners who could be located. Each of the 26 cases involved a suspect of a crime. In 9 cases, the examinee was indicted after the polygraph test because the charges were sufficiently corroborated by other material evidences. In 17 cases, the prosecution was dropped (non-prosecution) altogether either because suspicions were eventually cleared or the charges could not be corroborated at all. None of the selected cases was a stay-of-prosecution case, which means that the prosecution was suspended until new or critical evidence

could be secured.

### 2.2. Data

Two sets of scores were analyzed in each case. The first set consisted of the original scores for each case that were obtained during the investigation of a crime by the prosecutors' office. In each case, the original score was obtained by an official examiner who had at least 3 years of work experience after completing a training program. In each case, one chart was obtained for each of three repeated tests. Each chart was scored in accordance with the DoDPI criteria using the CC. Each chart contained a total of 8 scoring spots, one spot for each combination of four physiological channels (chest respiration, abdomen respiration, blood pressure, electro-galvanic skin response) and 2 relevant questions. Each case (examinee) yielded 24 spot scores (2 relevant questions x 4 channels x 3 charts) and the sum of those spot scores was the overall (grand total) score of the polygraph test.

Each spot was scored by assigning to it an integer value in the range of -3 and +3. For each relevant question, the score was given a negative or positive value, respectively, if the corresponding signal was

judged to depict a deceptive or a truthful response to the relevant question. The absolute value of the score represented the strength of the deceptive or truthful response.

An additional set of 182 cases were selected randomly from the archives of the Prosecutor's Office. This set of data was used to establish the standard against which the accuracy of the two sets of chart scores could be evaluated. The standard was developed by means of a Latent Class Analysis model based on the additional set of data.

### 2.3. Procedure

For the second set of data in the present study, each original chart was broken into four segments, one segment for each of the four physiological signal channels. Each segment contained a series of signals flowing from a single channel with two scoring spots corresponding to the responses to the two relevant questions. A randomized set of 312 segments (3 charts x 4 channels x 26 cases) was distributed to the examiners who scored the original charts. Each examiner rescored each scoring spots according to the DoDPI criteria. However, the CC could not be used because the examiners had no information regarding the remaining channels of the same case.

#### 2.3.1. Score Sets

Each set of scores consisted of the same 26 cases. The reliability and the accuracy were compared between the original chart scores and the segmented chart scores assigned by the same examiner.

## 3. Results

### 3.1. Descriptive Statistics of the Score Sets

Table 1 shows descriptive statistics, means and standard deviations, of the two score sets. At the time of the original polygraph tests, the suspects were indicted in 9 cases. Among the indicted cases, the examiner judged

that the suspect was telling a lie in 8 cases and telling the truth in 1 case. The prosecutor's office dropped 17 cases. Among the dropped cases, the examiner judged the suspect telling a lie in 8 cases and telling the truth in 4 cases at the time of the original polygraph tests. In 5 of the dropped cases, the polygraph test was inconclusive. A general trend to be noticed in the two score sets is that the examiners tended to be more conservative in rescored without the CC. That is, the segmented chart scores were generally higher than the original chart scores.

For the segmented chart scores, each of the 15 examiners scored the charts from all of the 26 cases. Therefore, it was possible to evaluate the degree to which different examiners agree with one another in scoring the same cases. The intra-class correlation coefficient based on the total score of each case from the segmented set of scores was 0.99. The 15 examiners showed almost perfect agreement in scoring the segmented charts.

### 3.2. Use of Consistency Criterion

Variation among spots scores within a case should be reduced by the use of the CC in scoring. To verify the use of the CC in the original scoring, a one-way ANOVA was performed with "Cases" as an independent variable of 26 levels and the 24 spot scores as dependent observations within each case. Although the "observations" for this analysis were not independent on one another, the purpose of this analysis was not to test the differences among cases but to determine whether the CC was actually used in the original scoring by comparing the variations in the two sets of chart scores. For the original chart scores, the values of the within-case sum of squares and between-cases sum of squares were 193 and 98, respectively. For the segmented chart scores, the values were 240 and 105. Based on the Levene test for homogeneity of variance,<sup>3)</sup>

3) Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford, California: Stanford University Press.

Table 2. Posterior Probabilities and Allocated Latent Classes of Indicator Patterns

Indicator Pattern		n	Posterior Probabilities		Allocated Latent Class
Indictment	Polygraph Decision		Deceptive Latent Class	Truthful Latent Class	
Indicted	DI*	60	1.00	0.00	Deceptive
Dropped	DI*	11	0.98	0.02	Deceptive
Stay-of-Prosecution	DI*	8	1.00	0.00	Deceptive
Indicted	NDI**	11	0.98	0.02	Deceptive
Dropped	NDI**	50	0.04	0.96	Truthful
Stay-of-Prosecution	NDI**	3	0.48	0.52	Truthful
Indicted	Inconclusive	18	1.00	0.00	Deceptive
Dropped	Inconclusive	19	0.21	0.79	Truthful
Stay-of-Prosecution	Inconclusive	2	0.86	0.14	Deceptive

\*DI: deception identified; \*\*NDI: no deception identified.

the original chart scores and the segmented chart scores differed significantly with respect to the within-case mean squares ( $t=4.03, p<.01$ ) but not the between-cases mean squares. These results were consistent with the hypothesis that the examiners used the CC in their original scoring of the polygraph charts.

### 3.3. Reliability

A measure of internal consistency was computed for each of the two score sets. For each case, the spot scores were treated as if they were responses to 24 items of a test. For the original chart scores and the segmented chart scores, Cronbach's  $\alpha$  was 0.93 and 0.91, respectively. That is, the internal consistency (degree of co-variation) of the 24 spot scores was almost equivalent for the two types of chart scores.

Because each examinee was tested three times with the identical set of questions, a subtotal score for each successive test can be computed by summing the 8 spot scores per test. Test-retest reliability then can be evaluated by comparing the three subtotal scores. Intra-class correlation among the three subtotal scores was 0.92 for the original chart scores and 0.84 for the segmented chart scores. Thus, the test-retest reliability of the original chart scores was higher than that of the segmented chart scores.

### 3.4. Accuracy

To establish a standard against which the accuracy of the two sets of chart scores could be evaluated, another set of 182 cases were selected randomly from the archives of the Prosecutor's Office. This new set of 182 did not include any of the original 26 cases.

#### 3.4.1. Establishment of the Standard

Following Albert et al.'s<sup>4)</sup> approach, a Latent Class Analysis (LCA) model<sup>5)6)</sup> with two latent classes, Deceptive and Truthful, was estimated based on the new set of 182 cases. For the LCA model, two nominal variables, polygraph examiner's decision (whether or not the polygraph examiner decided that the examinee was lying) and indictment (whether or not the examinee was eventually indicted), were used as manifest indicators of the latent classes. Since polygraph test is not admitted as

- 4) Albert, P. S., McShane, L. M., & Shih, J. H. (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*, 57, 610-619.
- 5) Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent Structure Analysis*, Boston: Houghton Mifflin.
- 6) Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.). *Applied Latent Class Analysis*. Cambridge: Cambridge University.

Table 3. Proportions of Decisions Suggested by the Original Chart Scores and the Segmented Chart Scores

Score Set	Latent Class	Decision Suggested by the Scores			Total
		DI*	NDI**	Inconclusive	
Original	Deceptive	16 (94.1%)	1 (5.9%)	0 (0%)	17 (100%)
	Truthful	0 (0%)	4 (44.4%)	5 (55.6%)	9 (100%)
	Total	16	5	5	26
Segmented	Deceptive	9 (52.9%)	3 (17.6%)	5 (29.4%)	17 (100%)
	Truthful	0 (0%)	7 (77.8%)	2 (22.2%)	9 (100%)
	Total	9	10	7	26

\*DI: deception identified; \*\*NDI: no deception identified.

evidence in court, prosecutors in Korea do not depend on the outcome of the polygraph test to make the decision of indictment. Therefore, the polygraph examiner's decision and the decision of indictment by a prosecutor are largely made independently with each other.

The examiner's decision had three categories: Deception Identified (DI), No Deception Identified (NDI), and Inconclusive. Indictment also had three categories: Indicted, Dropped, and Stay-of-Prosecution. The purpose of the LCA model was to identify indicator patterns (combinations of polygraph examiner's decision and indictment) that corresponded to the latent classes and to use those indicator patterns as the standard against which the accuracies of the original chart scores and the segmented chart scores could be compared.

To make the LCA model identified (i.e., to make the degrees of freedom of the model greater than zero), two constraints were imposed. First, the conditional probability that the examiner's decision was Inconclusive given the latent class of Deceptive and the conditional probability that the examiner's decision was Inconclusive given the latent class of Truthful were assumed to be equivalent. Second, the conditional probability that the examiner's decision was DI given the latent class of Truthful and the conditional probability of Indicted given the latent class of Truthful were assumed to be equivalent. The first assumption is based on a meta-analysis conducted on field studies by the Office of Technology Assessment of the U.S. Congress (1983, p. 52, Table 4)<sup>7)</sup> showing that the probabilities of

inconclusive results are similar in truly deceptive cases and truly truthful cases. The second assumption is also based on the same literature showing that the false positives are relatively rare in polygraph tests for real subjects (p. 52, Table 4) and on the general trend that 'not guilty' verdicts are similarly rare in criminal trials in Korea (2%-10% depending on the denominator, Supreme Prosecutors' Office of Korea, 2008).<sup>8)</sup>

The goodness of fit of the LCA model was excellent (likelihood ratio  $\chi^2=0.63$ ,  $df=1$ , ns). The probabilities of the latent classes were estimated to be 0.64 and 0.36 respectively for the Deceptive latent class and the Truthful latent class. Based on the posterior probabilities of the indicator patterns, 110 cases (60.4%) and 72 cases (39.6%) were allocated to the Deceptive latent class and the Truthful latent class, respectively (Table 2). Of 182 cases, 95.56% were estimated to be allocated to the latent classes correctly.

### 3.4.2. Comparison of Accuracies

Indicator patterns allocated to the latent classes by the LCA model were used as the standard against which the two sets of chart scores could be compared for accuracy.

7) U.S. Congress, Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation - A Technical Memorandum*. OTA-TM-H-15, November, Washington D.C.: U.S. Government Printing Office.

8) Supreme Prosecutors' Office of Korea (2008). *Comprehensive Examination and Analysis for the Year 2007 (Yr 2007 Jonghap Simsa Bunseok)*. Seoul: Supreme Prosecutors' Office.

A suspect was considered to have been truly truthful if (a) the prosecutor eventually dropped the case and (b) the examiner's original decision was either NDI or Inconclusive. In all other cases, the suspect was considered to have been truly deceptive.

To make a final decision, polygraph examiners in the Prosecutors' Office of ROK use two threshold cut-off points based on the recommendation of Abrams (1989).<sup>9)</sup> The examiner may decide that the suspect is lying (DI) if the grand total score, which is the sum of 24 spot scores in three repeated tests, is lower than -12. The examiner may decide that the case is inconclusive if the total score has a value of -12 to 0, inclusive. The examiner may decide that the suspect is telling the truth (NDI) if the total score is greater than 0. By using the same cut-off points and the standard established by the LCA model with a separate set of cases, the accuracies of the original chart scores and the segmented chart scores were estimated (Table 3).

The proportions of accurate decisions were 76.9% (20/26) and 61.5% (16/26) for the original chart scores and the segmented chart scores respectively. The odds of accurate decision by the original chart scores (0.769/0.231=3.329) were about twice as large as the odds of accurate decision by the segmented chart scores (0.615/0.385=1.597). When the Inconclusive category of the decision was combined into the NDI category, the proportions of accurate decisions were 96.2% (25/26) and 69.2% (18/26) for the original chart scores and the segmented chart scores respectively. The odds of accurate decision by the original chart scores (0.962/0.038=25.315) were about twelve times as large as the odds of accurate decision by the segmented chart scores (0.692/0.308=2.247). Overall, the accuracy of the original chart scores was considerably higher than that of the segmented chart scores. Although these differences in the proportions of accurate decisions between the two sets of scores are not statistically significant due to the small sample size, the sizes of the effect (odds ratios) are large enough to warrant practical significance for crime investigations. The difference in accuracy between

the two sets of scores is pronounced mostly in the false negatives (erroneously suggesting a deceptive case as truthful) as shown in Table 3.

#### 4. Discussion and Conclusion

The CC improved the reliability and accuracy of scores. The criterion is based on two premises: (1) when a signal from a physiological channel is elevated because the suspect is lying, similarly elevated signals also emanate from other channels whereas (2) the effect of extraneous factors, such as normal test anxiety, on physiological signals is random. The results from the present study suggest that these two premises are also likely to be valid because the scores assigned with the CC cannot be more reliable and valid than the scores assigned without the criterion when the premises are untrue.

The estimates of accuracy presented in this study are based on the assumption that the outcome of a polygraph test is determined solely by the grand total score of the test. However, in practice, the polygraph examiner makes the final decision by considering additional factors such as attempts of the examinee to take countermeasures. Therefore, the rates of accuracy estimated in the present study reflect the validity of the examiner's scores but not necessarily the validity of an examiner's final decision.

The examiners who participated in the present study were experts who had many years of experience using the consistency criterion to score the results of a polygraph test. When they were prohibited from using the criterion for rescoreing, some of them seemed to become more cautious in assigning scores possibly in an attempt to avoid a false positive decision error. This tendency may have resulted in generally higher scores on the segmented charts (Table 1) and may also have caused the differences in the reliabilities and the accuracies of the two sets of scores. Seasoned examiners without an experience of the CC might assign scores to the signals differently and in that case, the effects of the criterion on the reliability and the accuracy of polygraph scores might be different from those found in the present study. This possibility could not be examined in the present study because such scorers are not found in

9) Abrams, S. (1989). *The Complete Polygraph Handbook*. Lexington, MA: Lexington Books.

Korea. Worldwide research collaboration is needed to examine this possibility.

The absence of statistical significance of the difference in accuracy due to the small sample size should not be interpreted to mean that the CC is of little value. The direction of the difference very strongly suggests the opposite interpretation. Based on the results from the present study, continued use of the Consistency criterion for scoring polygraph charts can be recommended to detect lies in crime investigations.

In the present study, a standard was developed to measure the validity of test scores based on a combination of indictment status and polygraph test decision. In field studies to determine the absolute validity of polygraph test for crime investigation, confession is often considered as the best standard for establishing ground truth<sup>10</sup>). The presence or absence of confession, however, is often dependent on the outcome of polygraph test<sup>11</sup>). Without a gold standard, the accuracies reported in the present study do not indicate the absolute validity of the polygraph test used in the Prosecutors' Office of Korea. The standard used in the present study was to compare the relative superiority of different scoring methods but not to estimate the absolute accuracy of polygraph tests. Although the continued use of the consistency criterion for scoring is recommended, the present study is indifferent about whether polygraph test should be used to detect a lie in crime investigation.

## References

- Abrams, S. (1989). *The Complete Polygraph Handbook*. Lexington, MA: Lexington Books.
- Albert, P. S., McShane, L. M., & Shih, J. H. (2001). Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*, 57, 610-619.
- Ben-Shakar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 103-126). San Diego: Academic Press.
- Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenars & A. L. McCutcheon (Eds.). *Applied Latent Class Analysis*. Cambridge: Cambridge University.
- Iacono, W. G. (1991). Forensic "lie detection": Procedures without scientific basis. *Journal of Forensic Psychology Practice*, 1, 75-86.
- Iacono, W. G. & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinions. *Journal of Applied Psychology*, 82, 426-433.
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent Structure Analysis*, Boston: Houghton Mifflin.
- Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford, California: Stanford University Press.
- Raskin, D. C. & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 1-47). San Diego: Academic Press.
- Supreme Prosecutors' Office of Korea (2008). *Comprehensive Examination and Analysis for the Year 2007 (Yr 2007 Jonghap Simsa Bunseok)*. Seoul: Supreme Prosecutors' Office.
- U.S. Congress, Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation—A Technical Memorandum*. OTA-TM-H-15, November, Washington D.C.: U.S. Government Printing Office.

10) Ben-Shakar, G. (2002). A critical review of the control questions test (CQT). In M. Kleiner (Ed.), *Handbook of Polygraph Testing* (pp. 103-126). San Diego: Academic Press.

11) Iacono, W. G. (1991). Forensic "lie detection": Procedures without scientific basis. *Journal of Forensic Psychology Practice*, 1, 75-86.

원고접수 : 09.11.10

수정접수 : 09.12.17

게재확정 : 09.12.19