

A comment on the study by Sterzer, G., and Elaad, E.,  
Validity of the Control Question Test in Two Levels of the Severity of Crimes.  
IDENTA, 85.<sup>1</sup>

Assessing Polygraph Accuracy: The Importance of Choosing an  
Evaluation Technique which is Compatible with the Way the  
Examinations were Conducted.<sup>2</sup>

Avital Ginton,

Criminal Identification and Forensic Sciences Division, Behavioral Section, Israel National  
Police & Criminology Department, Bar-Ilan University, Ramat-Gan, Israel

#### Abstract

Sterzer and Elaad in their polygraph field study on Comparison Question Test (IDENTA, 1985), found that the false positive (FP) error rate in severe crimes was almost double compared to minor crimes. This was attributed to the difference in the degree of threat and the perceived consequences in the two levels of crime severity. Another factor, however, might have contributed to the difference in FP error rates found in that study. During the year 1979, the Israel National Police Polygraph Laboratory went through a substantial change in the way the polygraph examinations were conducted and evaluated. Unfortunately, this was ignored in the study, and the present analysis indicates that a great portion the crime severity effect might be an artifact of this change. On a more general level that goes beyond the present study, this case demonstrates the importance of choosing an evaluation technique which is compatible with the way the examinations were conducted.

Sterzer and Elaad in their study on Comparison Question Test (IDENTA, 1985) found that with numerical scoring technique, using a zero cutoff point with no inconclusive zone, results in false positive (FP) error rate of 23.33%

---

<sup>1</sup> **Anti-Terrorism; Forensic Science; Psychology in Police Investigations.**

Boulder, Colo.: Westview Press; Jerusalem: Heiliger and Co., 1986, (proceeding of IDENTA 85. The international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel.

<sup>2</sup> The comment published herein is a bit more detailed version compared to the original one.

for a minor crime sample (MC) and 43.14% for a severe crime sample (SC). The difference between these two error rates, which was found to be statistically significant, was attributed to the difference in the degree of threats posed by the perceived consequences of the relevant questions in the two levels of crime severity. Another factor, however, might have contributed to the difference in FP error rates.

During the year 1979, the Israel National Police Polygraph Laboratory went through a substantial change in the way the polygraph examinations were conducted and evaluated. The basic change was moving from a considerable reliance on the behavioral symptoms of the examinee and global evaluation of the polygraph records (as recommended by certain schools of thought), towards a major reliance on the semi-objective numerical scoring technique, which had by then become widespread. It was soon found that the change in the way information from the polygraph examinations is considered and evaluated, brought about a change in the manner by which the examinations were conducted, including a tendency to stress, during the pre-test interview and between charts, the control questions more than had been done previously. This tendency resulted in a considerable increase of physiological reactions to control questions.

That was evident in the records of the innocent examinees from the SC sample. Eighteen examinations were conducted during 1977-8, before the change took place, and twenty-two after it, during 1980-84<sup>3</sup>. It was found that in the earlier group, 55.6% of the records received negative numerical scores, (i.e., the reactions to the relevant questions in those records were stronger than to the control ones) compared to only 31.8% in the later one. Using the normal approximation to the binomial distribution, it was found that the probability associated with this difference is equal to or less than 0.065, ( $Z=1.487$ , one tailed) which is very close to an acceptable level of statistical significance.

---

<sup>3</sup> The SC sample also included 11 tests of innocent examinees from 1979, the year in which the change took place.

In contrast, the MC sample was taken mostly (27 of 30 records) from examinations conducted in the 1980s (1980-4), and the percentage of negative scores for these 27 records was 18.5%.

It seems that in order to obtain an estimate for the effect of crime severity on FP error rates in this study, only examinations conducted in the 1980s should be considered. When this is done it was found that the difference between the MC and SC samples regarding the FP error rates – 18.5% and 31.8% respectively (with no inconclusive zone), is associated with the probability of equal to or less than 0.142 ( $Z=1.07$ , one tailed), which is far from any acceptable level of significance. It means that, although a tendency towards the predicted direction of difference in FP error rates was observed, there is still a chance of 14% that the observed effect is due to sample error or chance fluctuation.

Furthermore, as correctly mentioned by Sterzer and Elaad, in actuality, a safeguard against FP (as well as FN) is taking place by using an inconclusive zone for low score outcomes. For instance, it is recommended by Barland and Raskin (1975) to use cut scores of +/- 6 to make a call, and records that receive scores between plus 5 and minus 5 are deemed inconclusive.

Applying this rule to the 1980s sub-samples of innocent examinees yields FP error rates of only 11.7% (two errors in 17 conclusive records) and 8% (two errors in 24 conclusive records) for the SC and MC samples respectively. Thus, it seems that practically, the factor of crime severity has only little effect on the FP error rates, and in order to better explore its actual existence a much larger sample is needed.

The main reason for this comment goes beyond the present study. It demonstrates that one cannot estimate the accuracy of polygraph examinations by using an evaluation technique at odds with the way in which the examinations were conducted. Examinations whose conduct was geared to the global evaluation approach which includes also behavioral-symptoms might be unsuitable for analysis by the more objective numerical scoring technique, or even by blind global evaluation, which is not accompanied by

behavioral-symptoms analysis. A proper evaluation technique should be compatible with the way the original examinations were conducted. This point has been, unfortunately, overlooked in many studies in the past, leading to unreliable and conflicting results.

#### References:

Barland, G.H. and Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12, 321-330.

Sterezer, G. and Elaad, E. (1985). Validity of the control question test in two levels of the severity of crimes. In: Anti-Terrorism; Forensic Science; Psychology in Police Investigations. Proceeding of IDENTA 85. The international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel. Boulder, Colo.:Westview Press; Jerusalem: Heiliger and Co., 1986,