

Assessing Polygraph Accuracy; The Importance of Choosing an Evaluation Technique Which Is Compatible With the Way the Examinations Were Conducted

Avital Ginton

AV-GN Polygraph, Applications & Research, Israel

Abstract

Sterzer and Elaad in their Israel National Police polygraph field study on Comparison Question Test (Sterzer & Elaad, 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 current 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 that 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% 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 about 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 comparison (control) questions more than had been done previously. This tendency resulted

1 This short article is an elaboration of a comment made on a study presented in IDENTA 85 by Sterzer & Elaad: Validity of the control question test in two levels of the severity of crimes. The original study and the comment were published 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.

2 At the time the original comment was made, Dr. Ginton was heading the Behavior Section in the Criminal Identification and Forensic Sciences Division of Israel National Police & a Faculty member in the Criminology Department, Bar-Ilan University, Ramat-Gan, Israel.



in considerable increased physiological reactions to those 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 1980-84³. 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.

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, UTAH approach recommended using cut scores of +/- 6 to make a

call, and records that receive scores between plus 5 and minus 5 are deemed inconclusive (Barland and Raskin, 1975; Raskin and Honts, 2002).

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 at most a small effect upon 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 study being discussed. It demonstrates that one could not get a fair estimation of 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 towards the global evaluation approach which includes for instance also the identification of developing patterns and trends along the repeated charts and behavioral symptoms, might be unsuitable for analysis by the more objective and strict numerical scoring technique, or even by blind global evaluation, which is not accompanied by behavioral-symptoms analysis. Some support for this conclusion can be drawn from a validity study which was conducted in the Israeli Police in 1979 and published a bit later (Ginton, Daie, Elaad & Ben-Shakhar, 1982). In that study, the polygraph tests and their analysis took place during the beginning of the transient period mentioned above concerning the manner by which the examinations were conducted and evaluated. Results of that study indicated that using numerical scoring was somewhat inferior compared to the global evaluation which was more compatible with the way the tests were conducted. Unfortunately, the point of compatibility was not addressed in the original study. This notion of compatibility considerations might be relevant in quite a number of other situations. For instance, two charac-

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



teristics of the Utah technique are 1) to review and discuss in a certain manner the questions between repeated charts and 2) in the scoring stage of a single-issue sequence, the Relevant questions are compared only to their adjacent preceding Comparison question (Raskin & Kircher, 2014). Unlike that, the Federal ZCT recommends as a default: 1) not to discuss the questions between charts and 2) in scoring the test (You-Phase) to pick for comparison the strongest reaction from the two Comparison questions surrounding each Relevant question (Federal Examiner Handbook, 2011). Applying the Utah scoring technique in that respect, to a ZCT examination in which no inter charts discussion took place, might results

in an increased FP rate. In the same vein, applying the federal scoring method in that respect to an examination that was conducted with inter charts discussion might results in a FN increased rate. That is because the inter charts discussion in the manner suggested by Utah method tends to increase in particular the salience of the Comparison question.

Unfortunately, this compatibility issue has been, overlooked in many studies, leading to unreliable and conflicting results. Finally, it is recommended to assimilate and incorporate this insight not only into the research arena but also to the everyday practice.



References

- Barland, G.H. and Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12, 321-330.
- Federal Examiner Handbook, (2011). Federal psychophysiological detection of deception examiner handbook. Special Edition. *Polygraph*, 40, 1, 1-66.
- Ginton, A., Daie, N., Elaad, E. and Ben-Shakhar, G. (1982). A method for evaluating the use of polygraph in a real-life situation. *Journal of applied psychology*, 67, 131-137.
- Raskin, D.C. and Honts, C, (2002). The comparison question test, in: Kleiner, M. (ed.) *Handbook of Polygraph Testing*. (2002), pp. 1-48, Academic Press, San-Diego, California, USA.
- Raskin, D.C. and Kircher, J.C., (2014). Validity of polygraph techniques and decision methods. In: Raskin, D.C., Honts, R.C. and Kircher, J.C. (Eds). *Credibility Assessment Scientific Research and Applications*. (2014), Pp.65-132. Elsevier, Academic Press, San Diego, CA, USA,
- 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,

