A Validity Study of the Comparison Question Test Based on Internal Criterion in Field Paired Examinations[1]

Avital Ginton[2]

Abstract

An estimate of the validity of polygraph field examinations was developed using a new strategy. Forty-two pairs of real life cases (mostly unverified), each consisting of two examinees that presented opposite versions regarding certain clear-cut points, were used. The 84 polygraph examinations were scored by experienced examiners. The examiners were completely blind as to the cases involved, and were not aware of their paired nature. Assuming that in each pair there must be one deceptive and one truthful subject, it was clear that each pair, which had been scored either as two deceptive or two truthful subjects, must include an error. Also, it was clear that a pair, which had been scored as one deceptive and one truthful subject, means either two hits or two errors. Based on these assumptions the error rate was computed. Results indicated an error rate of 19.5%. The problem of generalizing these results to polygraph field examinations as a whole is discussed.

Validity studies of polygraph examinations are basically divided into two categories. Either they are based on real life situations and examinations (field studies), or on some sort of laboratory experimental setting (analog studies).

The two categories suffer from different methodological weaknesses. In field studies it is hard to obtain a reliable criterion against which the polygraph results can be validated and to avoid a substantial sampling bias (Orne, 1975; Ginton, Daie, Elaad, & Ben Shakhar, 1982). While these weaknesses are usually controlled in analog studies, it is the lack of realistic fear of failure on the test which questions the usefulness of this type of studies for assessing the validity of polygraph examinations in real criminal investigations (Orne, 1975; Ginton et al. 1982).

Overcoming these methodological handicaps is almost mission-impossible within the conventional approach (but see Ginton et al. 1982), especially if one considers the limitations presented by ethical standards. In order to achieve this goal, all examinations must be perceived by the examinees as real life exams, and to a lesser degree, this applies also to the examiners' perspective. Clearly it points to real field examinations. By the same token, from the researcher's perspective the ground truth must be fully established, independent of the polygraph results and their effects, and without any sampling bias. That means examinations conducted under good laboratory set. Within the conventional approach it sounds like an "oxymoron," unless some problematic steps, ethically wise, are taken place. For instance, running field polygraph tests when there is no real need to find out whether the examinee is telling the truth or lies, because that has already been established by other means. Another example could be to deceive the examinees to believe that they are taking a real polygraph exam while in effect, it was but a laboratory research, without getting in advance their consent to participate in the research. (Ginton et al., 1982)

[1]Based on the same research a similar but not identical article was published in *Anti-Terrorism; Forensic Science; Psychology in Police Investigations*. Boulder, CO: Westview Press; Jerusalem: Heiliger and Co., 1986, Pp. 148-154 (Proceedings of Identa 85. The international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel).

[2]At the time the research was conducted (1985), Dr Avital Ginton, directed all the polygraph activities in Israel National Police, to include Polygraph Field Labs, Polygraph School, and Research Unit, as well as being a lecturer in the Department of Criminology in Bar-Ilan University, Ramat-Gan, Israel.

The present study tries to reduce the above-mentioned pitfalls by adopting a completely new approach at the time. It took advantage of the fact that a policy of using paired testing procedure on a regular basis has been applied in Israel Police Polygraph Laboratories since the early 80s and probably earlier.[3]

The study is based on field examinations conducted in cases where two opposing versions regarding the occurrence of specific events were presented by two examinees in each case. Selecting only the cases in which it was practically impossible to assume that the two parties were telling their subjective truth or that the two of them were lying, it was possible to estimate the validity of the tests by mathematical computations based upon the proportions of the various results obtained by the polygraph examinations per-se. Thus there was no need to rely upon confessions, convictions or any other sort of so-called verifications, which usually result in a substantial sampling bias.

Method

Eighty-four records of probable-lie comparison question polygraph examinations that had been conducted in the two main polygraph laboratories of the Israel National Police, during the years 1982-1984 were selected for this study.[4] The 84 records refer to 42 cases with a pair of examinations per case. Each pair consisted of one examinee that put some sort of blame on the other, who totally denied it, accusing the first examinee in fabricating the accusation. These 42 cases were the only cases that survived a selection procedure in which all paired cases from that pool (about 200), went through a careful content analysis of all the details involved in the controversies. The content analyses were performed by two experienced polygraph experts. These experts who got access to the detailed written documents explaining the case, which had been presented at the time to the polygraph labs, by the investigation units when asking for polygraph examinations, and later on to the original examiners before starting the tests.[5] Only cases in which both referees agreed that it was almost impossible to expect the two parties to be either both deceptive or both telling the truth regarding the relevant questions, were included in the research. The fact that the referees found only about 20% of the paired cases met the selection criterion indicates that one can trust that practically no case of two truthful examinees or two liars could be expected to fit ground truth.

There were 11 cases of minor sexual offense, two cases of theft, six cases of fraud, seven cases of police brutality and 16 cases of miscellaneous issues.

The 84 records were numerically scored by six experienced examiners using a seven-position scale per comparison, ranging from +3 to –3. Each record was scored by two examiners and using the Latin square design, each examiner scored a total of 28 records. The scorers were completely blind as to the cases involved and could not identify the pairs.

For detailed explanation of the numerical scoring technique the reader is referred to Barland and Raskin, 1975. Nevertheless, it is important to mention that in this technique each record receives a final numerical score. A negative score signifies that the autonomic responses to the relevant questions were on the whole stronger than to

[3]A similar procedure known as the "Marin Protocol" has recently been adopted by the APA for evidentiary polygraph applications. (APA Model Policy, 2007; Krapohl, 2005; Marin, 2000). In the Israeli police, since it is not allowed to introduce polygraph results as evidence in criminal cases, that policy has served only the investigation process (Ginton & Zoltak, 1991).

[4]The total number of examinations conducted at these two laboratories during 1982-4 was about 3500. Of them there were almost 200 paired cases with 400 examinations. The exact number is unavailable now due to technical reasons, but during those years between 8%-12% of the total volume were paired examinations.

[5]While doing the selection the referees did not have access to the polygraph outcomes or to the final outcomes of the investigations, however it might occurred that they were personally involved in some of the examinations at the time of their conductance.

the comparison questions, thus indicating deception to the relevant issue. On the other hand, positive final scores indicate truthfulness. Unlike in the common numerical scoring techniques, which usually contain an inconclusive zone for low final scores that are close to zero, in the present study, any final score other than zero has been considered an indicative score. However, records for which the two scorers gave opposite final scores were considered not indicative and deemed inconclusive. That applied also to records with two zeroes scores.

In case of 100% accuracy, the outcomes of each pair should be one positive and one negative score (opposite outcomes). However since we might expect errors, some pairs might produce, either two positive or two negative scores (identical outcomes).

Mathematical Analysis

Let us define the probability of having a correct decision as P(correct)=X, and the probability of having an incorrect decision as P(incorrect)=Y. Not allowing inconclusive results, a decision must be made whether correct or wrong. In probabilistic terms it means that P(correct) +P(incorrect)=X+Y=1.

Assuming that the examinations of the two parties in each pair were conducted independently, the probability of having correct decisions for both parties is equal to X^2. Following the same logic it is clear that the probability of having incorrect decisions for both parties is equal to Y^2, while 2XY expresses the probability of having one correct and one incorrect decision (see Table 1).

Table 1. A model of the distribution of outcomes in paired examinations of opposing versions. X=P(correct decision) ; Y=P(incorrect decision).

		Examinees presenting one side of the case		
		Probability of Correct Decision	Probability of Incorrect Decision	Total
Examinees presenting the counter side of each case	Probability of Correct Decision	X^2	XY	X
	Probability of Incorrect Decision	XY	Y^2	Y
	Total	X	Y	1.0

The percentage of pairs that had two identical outcomes (i.e., two positive or two negative scores per pair) is equal to the probability of having one correct and one incorrect decision (2XY). Similarly, the percentage of pairs that had two opposite outcomes (i.e., one positive and one negative score per pair) may be regarded as the joint probability of having either two correct or two incorrect decisions per pair (x^2+ y^2).

Hence given the percentage of identical and opposite outcomes it is possible to compute the values of X and Y which indicate the accuracy rate of detection of deception by polygraph examinations, by solving algebraic equations with two unknowns.

Results

As mentioned above, each record was scored blindly by two examiners. The correlation between the two scores with regard to the final decisions over the 84 records, which indicates the reliability, was

found to be 0.85[6]. Of the 84 records there were seven records (8.33%) from seven different pairs for which the two scorers gave opposite final scores. These records, which were not indicative, were considered inconclusive, and for the purpose of the present study they were eliminated from further computations. Thus, only 35 pairs (70 records) were left. The combined final scores of the two scorers for each of records were considered the joint final score of that record. Since in no case the two scorers gave zero scores, all the records have received either positive or negative joint final score.

In 24 pairs the two parties received opposite scores (i.e., one positive and one negative) and in 11 pairs the scores were identical in terms of direction (three pairs with positive scores for both sides, and eight pairs with negative scores). It should be mentioned that the very fact that only 11 of the 35 pairs were with Identical Outcomes, by itself, indicates that the tests' outcomes are not a matter of chance but probably indicates a valid procedure (p=0.04, Binomial exact test of significance, 2-tail). But in order to get a quantitative estimation of the accuracy, the algebraic computation took place to follow the aforementioned model.

The above data lead to the following equations:

Proportion of Opposite Outcomes: 24/35

$$X^2 + Y^2 = 24/35 = 0.686$$

Proportion of Identical Outcomes: 11/35

$$2XY = 11/35 = 0.314$$

A simple algebraic computation of quadratic equations with two unknowns, brought about the following solution:

$$X = 0.805 \text{ and } Y = 0.195.$$

That is to say that the accuracy rate was found to be 80.5%, and the error rate 19.5%.

Inspection of the various types of offenses in this study revealed that cases of police brutality seemed to have contributed a disproportionate number of errors that approaches a significant difference (njj=0.063, Fisher Exact Probability Test, one tailed). When this category was eliminated, 35 pairs remained, of which six were regarded as inconclusive, leaving 29 pairs (58 polygraph records) for computation. The results for these 29 pairs were:

$$X^2 + Y^2 = 22/29 = 0.759$$
$$2XY = 7/29 = 0.241$$
$$X = 0.86; Y = 0.14$$

Although the statistic only approaches significance, there are good reasons to believe that the difference in error rate of the police brutality cases reflects a real phenomenon. Alas, possible explanations for the higher proportion of errors in the cases of police brutality are beyond the scope of the present article, and will be presented elsewhere.

Discussion

The internal validity of the accuracy rate found in this study (80.5%) depends heavily on the assumption that the two examinations in each pair were conducted independently. Thus, one can raise the question that knowledge of the result of the first examination in each pair might influence the manner in which the second examination is conducted, leading to an artificial increase in the probability of obtaining an opposite outcome. This is especially acute when the two parties are tested by the same examiner. While it is impossible to ignore this danger, there are some indications that this is not the case. In the present study 20 pairs of examinations were originally conducted by two different examiners per pair, and only 22 pairs were conducted by one examiner per pair. Table 2 gives the distribution of the blind evaluation outcomes in these two categories. It is clear that there is not a substantial difference between the distributions of pairs of examinations conducted by a single examiner or by two different examiners (Excluding the inconclusive outcomes the chi-square with

[6]Unfortunately, I lost the raw material that indicated the exact type of correlation that was computed, and in the original version of this article which was published in 1986, it was not mentioned, but based on a following study (Ginton,1988), it seems that it was a Tetrachoric correlation (Ferguson 1966).

Yates correction = 0.013, df=1, p=0.909). It should be mentioned that in most cases in which two examiners were involved, the examinations were actually conducted parallel in time, eliminating any possible influence of one outcome on the other.

Table 2. Distribution of outcomes in paired examinations conducted by a single examiner per pair or two different examiners.

	ONE EXAMINER	TWO EXAMINERS	TOTAL
OPPOSITE OUTCOMES	13	11	24
IDENTICAL OUTCOMES	5	6	11
INCONCLUSIVE OUTCOMES	4	3	7
TOTAL	22	*20*	42

Another important question which stems from the present study is the generalizability of the results to polygraph field examinations as a whole. It is usually accepted that a main factor in producing false positive results is the increased level of anxiety that most innocent examinees experience on the test. In this respect, I believe that in the paired cases, the very fact that the innocent party in each pair is confronted by another person's direct accusation, would result in an elevated level of his tension and anxiety compared to the usual situation in which a person is confronted with a circumstantial suspicion. Thus the present situation is prone to a higher rate of false positive errors than usual.

On the other hand if a guilty person is confronted with such a direct accusation there is a high probability of confession before a polygraph examination takes place. Thus the guilty persons who took these examinations (i.e. did' not confess before the test) have already proved their resistance to interrogation, and should probably be considered "better liars" on the average than the normal population, and therefore a higher rate of false negatives is also expected. In conclusion, I believe that the present result is an underestimation of the general accuracy of polygraph examinations as a whole.

Finally, it should be noted that among the 11 pairs with identical outcomes there were three false negatives and eight false positives, which can be used with caution as a gross estimation of the distribution of error types.[7]

[7]In a following study which was an elaboration of the present approach, this suggested difference between the two types of errors has gained a clear support (Ginton, 1988).

References

APA Model Policy, (2007). Model policy for paired testing. *Polygraph*, 36 (2), 117-120.

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

Ferguson, G.A. (1966). *Statistical Analysis in Psychology and Education.* McGraw-Hill, New York.

Ginton, A. (1986). A built-in validity in polygraph field examinations. *Anti-Terrorism, Forensic Science & Psychology in Police Investigations.* Boulder, CO : Westview Press; Jerusalem: Heiliger and Co., pp. 148-154 (Proceedings of Identa 85. The international congress on techniques for criminal identification & counter terrorism. Jerusalem, Israel, 1985.

Ginton, A. (1988). Estimating accuracy of the control-questions-test, using a self validating set of real-life polygraph examinations. Paper presented at the "NATO Advanced Study Institute of Credibility Assessment," Maratea, Italy, June 14-21, 1988.

Ginton, A., Daie, N., Elaad, E., & Ben Shakhar, G. (1982). A method of evaluating the use of the polygraph in a real life situation. *Journal of Applied Psychology*, 67, 131-137.

Ginton, A. and Zoltak, S. (1991). *Internal Regulations and Guidance* (in Hebrew). Israel National Police, Scientific Interrogation Laboratory, Behavior Section, Division of Criminal Identification and Forensic Sciences. Written by then-Chief Superintendent Dr. Avital Ginton and then-Major Shmuel Zoltak.

Krapohl, D., (2005). Polygraph decision rules for evidentiary and paired-testing (Marin Protocol) applications. *Polygraph*, 34(3), 184-195.

Marin, J., (2000). He said/She said: Polygraph Evidence in Court. *Polygraph*, 29(4), 299-304.

Orne, M. T. (1975). Implications of laboratory research for the detection of deception. In: Ansley, N., (Ed.) *Legal Admissibility of the Polygraph*, Charles C. Thomas, Springfield, Illinois, U.S.A. pp. 94-119.