

evidence for validating lie detection tests would be gained from thorough background investigations. A confession of guilt (or the admission of an attempted deception) is often used for estimating validity but it is not a completely satisfactory independent criterion. It is rarely clear whether the confession came before the polygraph was attached or after; or whether a complete polygraph test was run; or whether the interrogator made his "judgment of deception based on the polygraph" before or after the confession. To put it simply, the nature of police or security work does not lend itself readily to precise experimental control. In many such cases, independent verification by other than self-incriminatory means, may not be achievable.

Lie detection would exhibit high validity when polygraph-derived data are consistent with independent data on deception, such as when those judged to be deceptive are later found to have been deceptive, and those judged to be non-deceptive are later found to be nondeceptive, etc. There would be low validity when those judged to be truthful are found later to have practiced deception; or when those judged to be deceptive are found later to have been truthful. In real life, the problem of determining validity is complicated because those who are judged to be deceptive are not ordinarily hired and that ends the matter; no further investigation is conducted to determine whether or not the person actually was deceptive, although that would be required to clarify the problem of validity. A thorough appraisal of validity would require data in nine cells:

Polygraph judgment	Independent evidence		
	Guilty	Innocent	Indeterminate
Guilty (or practicing deception).....	*		
Innocent (or not practicing deception).....		*	
Indeterminate (no conclusion possible).....			

Obviously, high validity would require that the preponderance of cases fall in the starred (*) cells; and validity would decrease as the percent increases in any of the other cells.

B. RELIABILITY

Reliability measures the extent to which a test produces consistent or reproducible results. Reliability refers to the accuracy of measurement and should not be confused with validity. A test cannot be valid without also being reliable. Various aspects of reliability can be measured in the following ways:

- (1) Comparing the results achieved by two or more examiners working independently on the same case material.
- (2) Comparing the results of two or more tests on the same person taken at separate but close time intervals.
- (3) Comparing the results of one part of an examination with another, e.g., odd versus even items on one subtest, two different physiological indexes, or two different methods of examinations (viz., peak of tension versus questionnaires).

In the current practice of lie detection, no attempt is made to measure the absolute values of the three physiological responses which are being recorded, though many such schemes have been proposed. The examiner judges the responses in a qualitative fashion, using visual inspection to compare the magnitude and pattern of responses to relevant and irrelevant questions. There is no objective method of reporting test results.

Little attention is directed to the accuracy of the three instruments used in the polygraph, though they may be precise enough for present purposes. The breathing and blood pressure instruments operate on pneumatic pressure and their response characteristics are obviously nonlinear. According to one manual,² air leaks in these two systems should not exceed a pen excursion of 0.25 inches in 30 seconds for the pneumograph and of 0.25 inches in 10 minutes for the sphygmomanometer. Disregarding the rate of leak, this is a 5-percent error over the entire scale of 5 inches; the true error would be two to three times larger than 5 percent because the three tracings share the 5-inch scale. The psychogalvanometer is a sensitive instrument which must be adjusted continuously to contain the responses on the scale; some units incorporate a self-centering feature. Darrow (1929) and Lacey (1949) have shown that, among the several possible ways of measuring the GSR responses (based on conductance or resistance), the log change in conductance is the most reliable one. Martin

² Prepared by the Office of Naval Intelligence.