
E31 Using Computer Measurements to Improve Evaluator Reliability of Credibility Assessment With Human Physiology Waveforms

James P. O'Burke, MS, The Polygraph Institute, San Antonio, TX 78258*

Learning Overview: After attending this presentation, attendees will understand how using computer measurements of human physiology can aid in evaluator decision-making using an analytical theory for credibility assessment. Historically, evaluators utilized visual identification of pattern analysis in human physiology waveforms to make determinations of truth or deception. This presentation will show how combining objective computer measures with data quality assessments can improve the reliability of decisions made by evaluators.

Impact on the Forensic Science Community: Credibility assessment is now 100 years old yet suffers from a mixed reputation of acceptance. This presentation will impact the forensic science community by increasing openness to embracing the potential accuracy of credibility assessment when data quality standards are utilized and evaluator objectivity is increased for decisions with computer measurement tools. Previously, credibility assessment standards have not provided clear data quality standards and have not incorporated computer measurements of data to aid evaluators in decision making.

In polygraphy, the Comparison Question Technique (CQT) is used to contrast physiological arousal during multiple presentations of relevant issue questions with comparison questions. Evaluators award diagnostic points at each relevant question and physiology component for an aggregate numerical test score. Exceeding recommended cut scores allows for test decisions. Previous studies have used small numbers of evaluators to analyze large numbers of exams to demonstrate efficacy of this methodology. Previous studies have demonstrated accuracy levels better than chance, yet also demonstrate a higher-than-expected Standard Deviation (SD) for aggregate test scores. It is unknown if this SD derives from differences in test data quality or from the evaluators in reliably assessing decision points during a test evaluation.

In this study, more than 100 evaluators were used to evaluate three field polygraph exams with confirmed results to better understand SD. Each field exam was a single issue with 2 relevant questions and 18 decision points that aggregated for a test score. This study treated each decision point as 54 unique decisions from the three exams to examine evaluator concurrence. Examiner concurrence is defined as at least 90% of the evaluators.

A Data Quality Coefficient (DQC) was also created for each of the 54 decision points, along with a computer measurement of the physiological waveform data collected to contrast with evaluator decisions. This method should identify if computer measures would aid evaluators in their decision-making, thus increasing inter-rater reliability. Further, this study should aid in examining the extent to which data quality contributes to SD and variance.

There were three confirmed polygraph examinations used for this study: one with a high aggregate decision score for deception, one with a moderate aggregate decision score for truthfulness, and one with an exceptionally low aggregate decision score for an inconclusive result. For the two conclusive exams, the evaluators concurred on the correct analysis using visual analytics. However, the evaluators only concurred on 14 of the 54 decision points contained in the three polygraph exams. As well, on the inconclusive exam, approximately 20% of the evaluators incorrectly made conclusive opinions of either truth or deception opinions on the low data quality exam. There was a direct correlation to examiners awarding diagnostic points at those decision points with high DQCs, with a reduction in diagnostic decision points on moderate and low DQCs. Most evaluators agree the easiest waveform for analysis is the Electrodermal Activity (EDA) component where this study had the highest evaluator concurrence at 68.6%. Arguably, the most complex component for visual analysis is the respiratory waveform. In this study, respiration had the lowest evaluator concurrence at 44.7%. The use of computer measurements of the respiratory component alone would have improved evaluator concurrence by 55.5% of those decision points and reduced several evaluator errors.

This study supports the idea that evaluators would benefit from data quality standards in making decisions. Currently, the complexity of assessing data quality is still uniquely best handled by human evaluators. However, this study clearly suggests evaluators would improve their reliability by using objective computer measurements to augment decision-making where computers demonstrate greater objectivity.

Pattern Analysis, Decision Reliability, Credibility Assessment