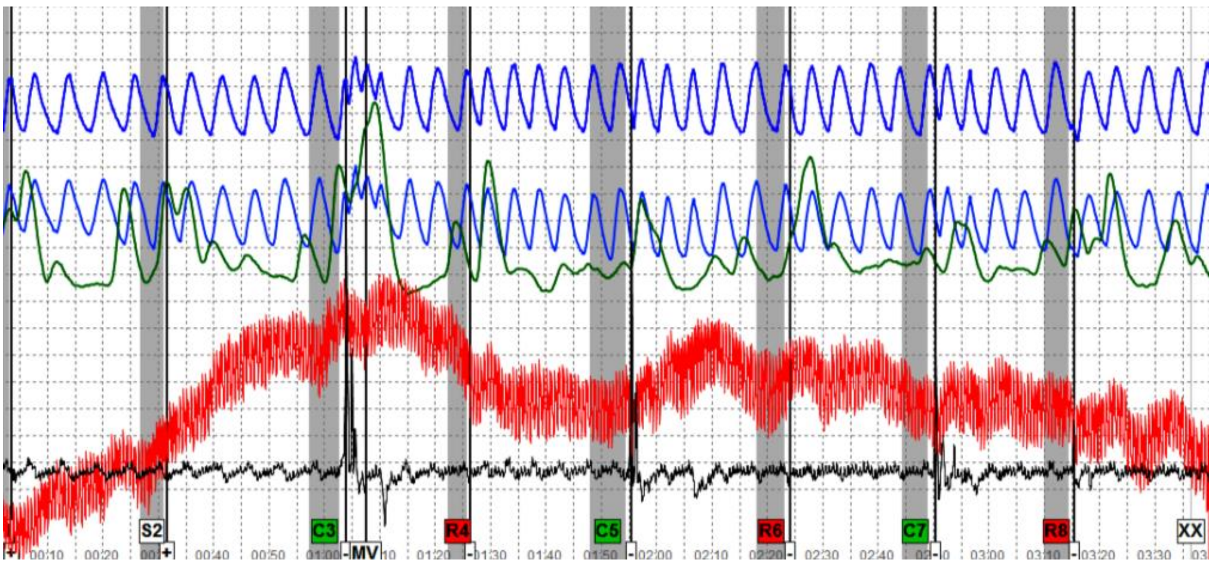
Tech Talk: Introduction to the Lafayette OSS-3 Test of Proportions to Discriminate Countermeasures and Random Artifacts

Raymond Nelson

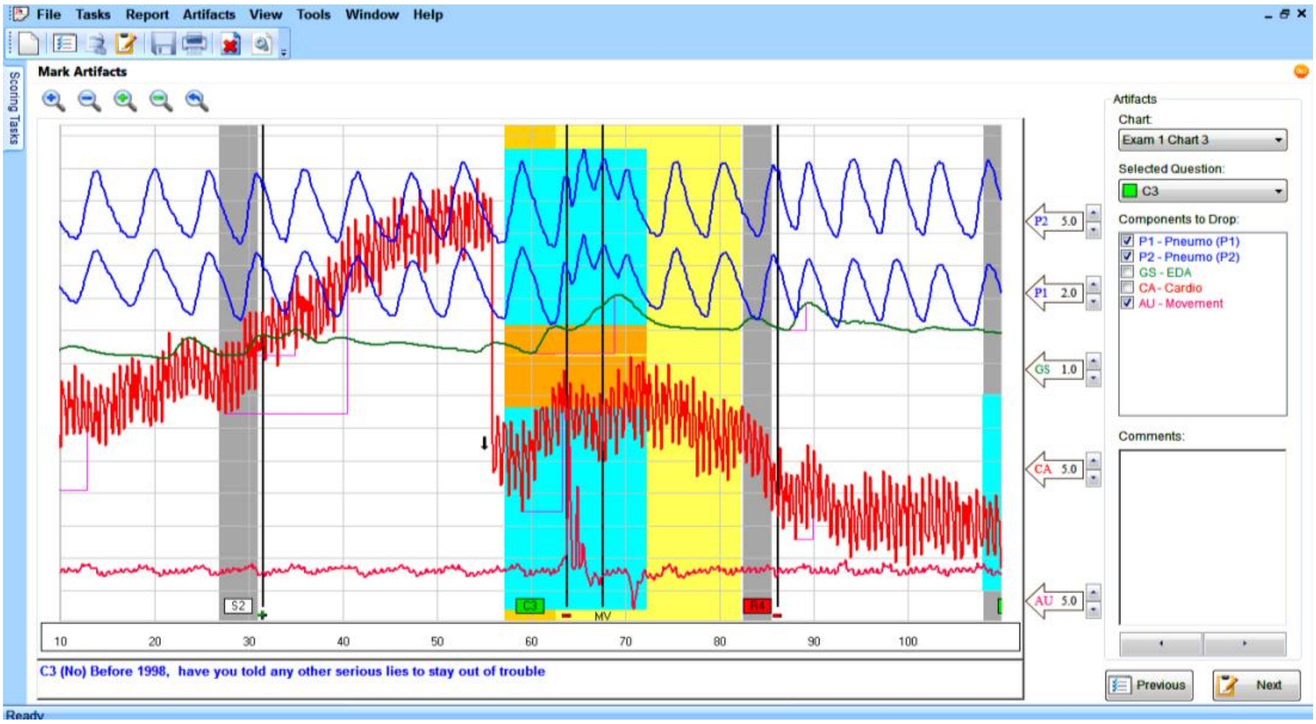
The Objective Scoring System version 3 (OSS-3) is a powerful computer algorithm that can calculate a probabilistic classifier for both diagnostic and screening polygraphs. The OSS-3 algorithm provides helpful options and features, one of which is the ability to mark artifacted segments that should not be included in the statistical analysis. Figure 1 shows a chart with conspicuous activity artifacts at the comparison questions and an obvious respiration artifact at the first comparison question.

Figure 1: Chart with Activity Artifacts



Most examiners will know intuitively when they observe consistent activity only at the comparison questions that it is most likely a systematic attack in attempt to alter the test result. Allowing the algorithm to score and interpret these artifacted responses is incorrect. Artifacted segments should be marked and excluded before allowing the algorithm to calculate a result. Fortunately, the Lafayette LXSoftware and OSS-3 scoring algorithm provide a simple to use interface to accomplish this. Open the exam PF, launch the OSS-3 algorithm by selecting the ***Series*** menu and then selecting the ***Score*** → ***OSS-3*** menu items. The examiner will need to provide the OSS-3 algorithm with information about the type of exam, and will also need to verify that the algorithm has selected the desired charts and questions, before clicking the ***Next*** button at the lower right side of the computer screen. The examiner will then have an opportunity to scroll through all selected charts and questions, to mark any observed artifacts. Use the ***left arrow*** **←** and ***right arrow*** **→**, or the onscreen arrow buttons, to scroll through the questions. Figure 2 shows the OSS-3 screen display while reviewing the data for artifacts.

Figure 2: On-screen artifact review



The *selected question* is highlighted in light yellow, and the *15-second evaluation window* is highlighted in gold. On the right of the screen is a list of sensors. Checking the box next to the sensor name will exclude the segment from the statistical analysis. Advance through the sequence of questions and mark any observed data artifacts. In addition to excluding marked artifact segments from the statistical analysis, the OSS-3 algorithm will also calculate a supplemental statistic, the ***test of proportions***.

The ***test of proportions*** is a simple and well known statistical test to compare binary outcomes in two groups. If the ***test of proportions*** is statistically significant (typically with alpha = .05) then the data can be said to support a scientific conclusion that the groups are different. The ***test of proportions*** can be used in the polygraph context to analyze the occurrence of data artifacts at relevant questions and other types of test questions. The ***test of proportions*** will use information about the size of the two groups, along with information about the observed frequency counts of each the two outcome possibilities for each group, together with some basic statistical theory to calculate the statistical result.

Before proceeding further, it is useful to remember that the analytic theory of the polygraph is that greater changes in physiological activity are loaded at different types of test stimuli (i.e., differential salience) as a function of deception or truth-telling in response to the investigation target stimuli. This theory can encompass CQT and CIT exams and explains both PLC and DLC exam formats. This analytic theory is limited to what we expect to observe in the data. It does not attempt the impossible task of defining exactly what an examinee may be thinking or feeling during testing. The basic idea behind any test is this: stimulus-and-response. Present the stimulus, then observe and quantify the response. Then aggregate and reduce the data and calculate the probabilistic and categorical test results. In the end, all test results and all scientific sion with the strength of evidence in support some other possible explanation for having observed a result.

It is reasonable in the stimulus-and-response testing paradigm to assume that an observed response was caused by the test stimulus under two conditions. First, the response must be timely with the stimulus. Second, there must be no other observable possible cause for the observed response. Data artifacts represent an alternative possible cause for an observed response. In the context of a data artifact, we do not know whether the observed response was caused by the test stimulus or the artifact event. Neither do we know the actual cause of observed data artifacts. This is okay, because a goal of science and scientific experiments is often to quantify things that are difficult to measure and that is where statistics are useful.

The ***test of proportions*** serves as a form of scientific experiment to test the hypothesis that the examinee is systematically attacking one type of test stimuli more than the other. The null-hypothesis in this experiment is that there is no difference and that the pattern of artifacts is characterized by random chance.

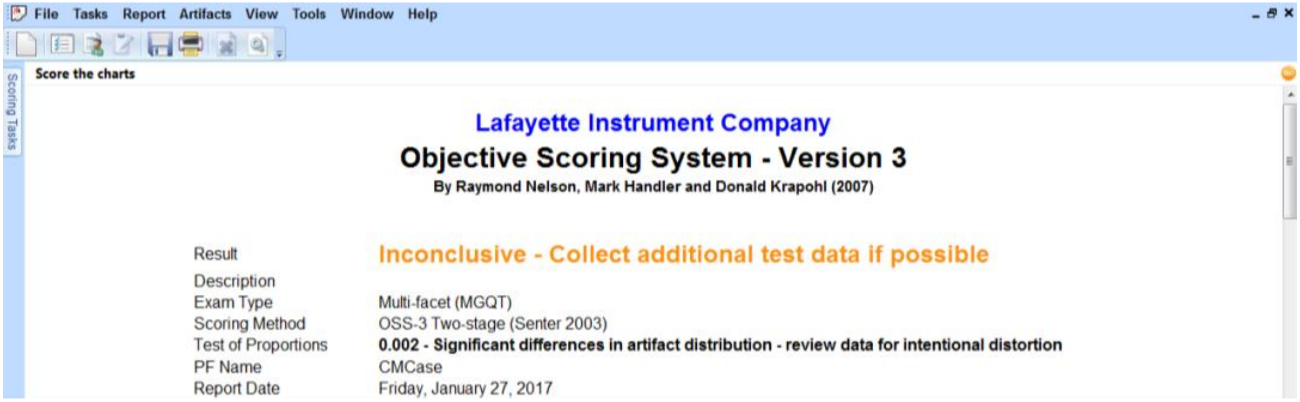
The ***test of proportions*** uses information about the number and location of observed artifacts to make a statistical calculation of the probability that the observed artifacts are consistent with what we would expect to observe if the artifacts were cause by random chance and not a systematic or strategic attempt to alter the polygraph test result. When the probability is sufficiently low ($p < a$), that the number and location of observed artifacts is random, then the data can be said to be supportive of a scientific conclusion that the observed artifacts are non-random and are instead a systematic attempt to fake or alter the test result – a countermeasure. The ***test of proportions*** relieves us from making subjective and unreliable decisions about what constitutes a consistent pattern of artifacts.

This works because countermeasures, to be effective, will have to consistently alter the valence (sign value) of the numerical transformations of the recorded test data from the *stimulus-and-response* trials. The examinee must do this in a controlled and skillful manner for which the voluntary activity are indistinguishable from normal autonomic nervous system activity (therein is the real difficulty). Voluntary activity that is likely to alter the valence of the feature extraction is also likely to result in data of artifacted and unusual quality. Finally, non-systematic or random attempts to alter the test result is unlikely to be successful reversing the valence of the resulting aggregated numerical scores, and will most likely result only in suspicions of non-cooperation and an inconclusive test result.

The published literature on countermeasures can be summarized succinctly as follows: 1) although members of the public have expressed an opinion that the polygraph is easy to defeat, studies show otherwise; 2) although polygraph examiners have made claims of expertise in detecting countermeasures, published studies have shown that examiner may not be successful at reliably discriminating countermeasures from random artifacts; and 3) the polygraph continues to discriminate deception and truth-telling at rates significantly greater than chance despite the fact that virtually every innocent person and every guilty person will do something in attempt to appear truthful during polygraph testing.

Observed artifacts that differ at a statistically significant level ($a = .05$) from random chance can be said to support a conclusion that the examinee has systematically attempted to alter the test result. Computation of the statistical result is a function of the number of presentations of each of the types of test stimuli, along with the number of artifacts observed at the different types of test stimuli and a normal approximation to the binomial distribution. Figure 3 shows the OSS-3 report header when the ***test of proportions*** is statistically significant. In this case, the test result is shown as inconclusive because the ***test of proportions*** is significant. In the event that the test result is statistically significant for deception, the examiner would be alerted to that result.

Figure 3: OSS-3 Report Header



Lafayette Instrument Company

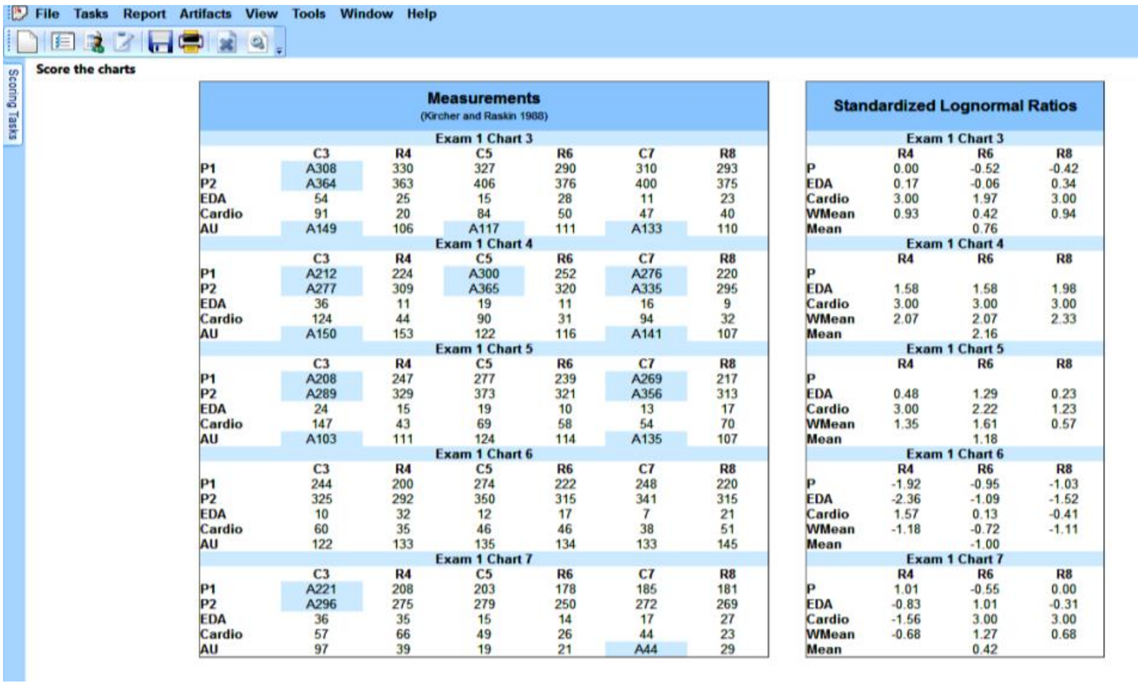
Objective Scoring System - Version 3

By Raymond Nelson, Mark Handler and Donald Krapohl (2007)

Result	Inconclusive - Collect additional test data if possible
Description	
Exam Type	Multi-facet (MGQT)
Scoring Method	OSS-3 Two-stage (Senter 2003)
Test of Proportions	**0.002 - Significant differences in artifact distribution - review data for intentional distortion**
PF Name	CMCase
Report Date	Friday, January 27, 2017

When the ***test of proportions*** is statistically significant the information will be highlighted in **bold** print. Reported information includes the level of statistical significance and a suggestion to review the data again before deciding whether the data do or do not support a conclusion that the examinee as attempted to falsify the test result by using countermeasures. Figure 4 shows the lower portion of the OSS-3 report including an audit of all data artifacts that were marked by the examiner, excluded from the statistical test result, and included in the ***test of proportions***. A Quality Assurance reviewer can easily refer to the recorded test data to confirm that the examiner has made correct and reasonable decisions in excluding the marked artifacted data segments.

Figure 4: Audited Artifact Marks



Measurements (Kircher and Raskin 1988)

	C3	R4	C5	R6	C7	R8
Exam 1 Chart 3						
P1	A308	330	327	290	310	293
P2	A364	363	406	376	400	375
EDA	54	25	15	28	11	23
Cardio	91	20	84	50	47	40
AU	A149	106	A117	111	A133	110
Exam 1 Chart 4	C3	R4	C5	R6	C7	R8
P1	A212	224	A300	252	A276	220
P2	A277	309	A365	320	A335	295
EDA	36	11	19	11	16	9
Cardio	124	44	90	31	94	32
AU	A150	153	122	116	A141	107
Exam 1 Chart 5	C3	R4	C5	R6	C7	R8
P1	A208	247	277	239	A269	217
P2	A289	329	373	321	A356	313
EDA	24	15	19	10	13	17
Cardio	147	43	69	58	54	70
AU	A103	111	124	114	A135	107
Exam 1 Chart 6	C3	R4	C5	R6	C7	R8
P1	244	200	274	222	248	220
P2	325	292	350	315	341	315
EDA	10	32	12	17	7	21
Cardio	60	35	46	46	38	51
AU	122	133	135	134	133	145
Exam 1 Chart 7	C3	R4	C5	R6	C7	R8
P1	A221	208	203	178	185	181
P2	A296	275	279	250	272	269
EDA	36	35	15	14	17	27
Cardio	57	66	49	26	44	23
AU	97	39	19	21	A44	29

Standardized Lognormal Ratios

	R4	R6	R8
Exam 1 Chart 3			
P	0.00	-0.52	-0.42
EDA	0.17	-0.06	0.34
Cardio	3.00	1.97	3.00
WMean	0.93	0.42	0.94
Mean		0.76	
Exam 1 Chart 4	R4	R6	R8
P			
EDA	1.58	1.58	1.98
Cardio	3.00	3.00	3.00
WMean	2.07	2.07	2.33
Mean		2.16	
Exam 1 Chart 5	R4	R6	R8
P			
EDA	0.48	1.29	0.23
Cardio	3.00	2.22	1.23
WMean	1.35	1.61	0.57
Mean		1.18	
Exam 1 Chart 6	R4	R6	R8
P	-1.92	-0.95	-1.03
EDA	-2.36	-1.09	-1.52
Cardio	1.57	0.13	-0.41
WMean	-1.18	-0.72	-1.11
Mean		-1.00	
Exam 1 Chart 7	R4	R6	R8
P	1.01	-0.55	0.00
EDA	-0.83	1.01	-0.31
Cardio	-1.56	3.00	3.00
WMean	-0.68	1.27	0.68
Mean		0.42	

All examinations may have some data artifacts. These data artifacts, along with the need to produce small errors of measurement, are reasons that tests commonly use multiple sensors, multiple test stimuli, and multiple repetitions of the sequence of test stimuli. When an examinee systematically attacks every instance of a specific type of stimuli, then it is an easy and intuitive conclusion that we are observing some strategic and deliberate activity. A more difficult scenario is when an examinee engages a strategy of inconsistency or feigned inconsistency. Remembering that a countermeasure is a systematic attempt to alter the test result, scientific conclusions about countermeasures should be based on an analysis of the data and probability theory.

Where data artifacts are non-systematic, they can be characterized as a random process for which we can use our familiar statistical distributions to characterize. In other words, we can use the ***test of proportions*** to calculate the probability that observed data artifacts do or do not fit what we would expect to observe if they were random. When artifacts are consistent with a random distribution of artifacts they will be much less likely to have any effect towards successfully reversing the cumulative valence of the numerical transformations from the stimulus-and-response trials. When the observed artifacts differ significantly from random, they can be expected to have the potential to alter the valence of the cumulative numerical scores. Ultimately both artifacts and countermeasures should be removed from the statistical analysis. Responses due to random artifacts will have no diagnostic value and can only weaken the statistical calculations concerning the test result. The Lafayette OSS-3 ***test of proportions*** give us a convenient and powerful tool to help us quantify our answers to questions about countermeasures.

The advantage of a statistical approach is that it provides us with a rational way to resolve those more ambiguous situations when we observe data artifacts at some, but not all, presentations of one of the stimulus types, and the even more ambiguous situation when we observe data artifacts at multiple types of test stimuli. A statistical approach to countermeasures replaces the impossible task of divination or mind-reading the strategic intent with an analytic process. More practically, a statistical approach serves to answer questions that scientists will eventually ask, regarding the quantifiable strength of evidence to support our conclusions, and the associated probabilities of error. The primary advantage of a statistical and analytic approach to the question of polygraph countermeasures is that a statistical approach moves the countermeasure discussion into the realm of reproducible analytic results that are expected of science and scientific testing. The next generation of computerized polygraph scoring algorithms will undoubtedly have even greater capabilities.