Multinomial Reference Distributions for the Empirical Scoring System1

Raymond Nelson2

Abstract

Scoring and interpretation of CQT data has progressed from subjective visual interpretation to the use of structured feature extraction methods and analytic models that make use of statistical decision methods. Empirical reference distributions are now available for a variety of comparison question polygraph test formats and numerical scoring methods. However, no previously published description could be found for a theoretical reference distribution for CQT scores. Theoretical reference distributions are an important aspect of all areas of science because, as the name suggests, they depend fundamentally on a coherent and practical understanding of the underlying theoretical basis such that it can be expressed mathematically. Theoretical distributions are calculated from facts or assumptions that are subject to logical mathematical proof. Theoretical distributions can be used to make inferences about empirical data, and can also be useful as a likelihood function for Bayesian analysis. An advantage of the theoretical distribution and a Bayesian approach is that the replacement or addition of evaluation features and recording sensors can be a simple matter when naïve assumptions are made. Multinomial reference distributions for CQT scores are calculated under the null hypothesis to the analytic theory of the polygraph and the CQT, and the results from closed form calculations were compared graphically to a Monte Carlo simulation. A description of the calculation of the multinomial reference distributions is provided for replication and for readers who wish to develop their understanding of, and intuition for, multinomial distributions. Reference tables for random discrete uniform multinomial distributions for the variety of CQT formats are provided in appendices.

Introduction

A combinatoric3 solution is described herein for the computation of multinomial4 statistical reference distributions5 for empirical scoring system6 (ESS) scores for comparison question test (CQT) data. Availability of a theoretical referenced distribution for the CQT can help to advance the science of the polygraph and credibility assessment testing through the comparison of real world observations with expected results as defined by a mathematical and statistical model. In addition to the availability of empirical data and empirical reference distributions, theoretical reference distributions help to understand the validity of an area of scientific theory, and can help to better understand and better interpret empirical observations and empirical data.

Use of statistical reference distributions to interpret polygraph data was first suggested by Barland (1985) who described a Gaussian Gaussian signal discrimination model (Wickens, 1991; 2002), though this was largely unnoticed until the introduction

of the Objective Scoring System (OSS; Krapohl & McManus, 1999; Krapohl, 2002) and the later Empirical Scoring System (ESS; Nelson, Krapohl & Handler, 2008; Nelson et al., 2011). Empirical reference distributions were published by Nelson and Handler (2015) for all comparison question polygraph formats for which data was included in the met analytic survey by the American Polygraph Association (2011). Although empirical reference distributions are becoming more widely used by polygraph field examiners in recent years, no published description exists for the calculation of a theoretical distribution for CQT scores.

Use of statistical reference distributions to interpret polygraph data was first suggested by Barland (1985) who described a Gaussian Gaussian signal discrimination model (Wickens, 1991; 2002), though this was largely unnoticed until the introduction of the Objective Scoring System (OSS; Krapohl & McManus, 1999; Krapohl, 2002) and the later Empirical Scoring System (ESS; Nelson, Krapohl & Handler, 2008; Nelson et al., 2011). Empirical reference distributions were published by Nelson and Handler (2015) for all comparison question polygraph formats for which data was included in the metaanalytic survey by the American Polygraph Association (2011). Although empirical reference distributions are becoming more widely used by polygraph field examiners in recent years, no published description exists for the calculation of a theoretical distribution for CQT scores.

---

1 This project was supported and made possible by the Lafayette Instrument Company where Raymond Nelson is employed as a research specialist.

2 Raymond Nelson is a research specialist with Lafayette Instrument Company, which develops and markets polygraph technologies. Mr. Nelson is a polygraph field examiner and psychotherapist with expertise in sexual offending, victimization, trauma and development in addition to other experience in testing, data analytics and statistics. Mr. Nelson is one of the developers of the OSS3 computer scoring algorithm and has published numerous studies on the ESS and other aspects of the polygraph. Mr. Nelson serves as an expert witness in legal matters involving both polygraph and psychology/ psychotherapy. Mr. Nelson is a past president, and currently elected member of the APA Board of Directors, and has helped with policy development at the state, local and national level. The views an opinion expressed herein are those of the author and not the APA or LIC.

3 Combinatorics is an area of mathematics that involves counting the combinations of objects that can be created from a defined set of items according to certain rules or constraints. A number of textbooks, such as the one by Skiena (1990) and Chen and Koh (1992) address this topic in detail.

4 Multinomial refers to a statistical distribution of the expected frequency of possible outcomes under repeated trials when there are multiple possible outcomes for each individual trial. Applied to the polygraph context each presentation of each test stimulus and each sensor score represents an individual trial for which the outcomes maybe coded in as + or 0. The more common binomial distribution, with two possible outcomes for each trial, is a special case of the multinomial. Detailed information can be found in mathematics texts and reference such as by Abramowitz & Stegun, (1972) and Olver, Lozier, Boisvert, & Clark (2010).

5 A distribution is a numerical and mathematical description of the range of possible values for a random variable. A random variable is a value that is unknown and can take a variety of possible values. Statistical distributions are mathematical or empirical descriptions of the range of values and the expected proportion or probability of observing each unique value if they occur due to random chance alone. More information can be found in statistics textbooks such as the by Evans, Hastings & Peacock. (2010) and Spiegel (1992).

6 The ESS is an evidencebased standardized protocol for the analysis of comparison question polygraph data, and is largely a derivative product of earlier research by others, including: Kircher and Raskin (1988), Raskin, Kircher, Honts and Horowitz (1988), Kircher, Krisjianssen, Gardner and Webb (2005), Krapohl and McManus (1999), and Senter and Dollins, (2003).

Statistical reference distributions7,8, are said to be theoretical then they are calculated from basic facts and assumptions that are accepted as the product of logical and mathematical proof. This is in contrast to empirical distributions that are calculated from observed sampling data9. In practice, theoretical and empirical reference distributions are often used together10. Part of the usefulness of mathematical/theoretical distributions is that probability statements about the statistical significance of observed data are mathematical abstractions that may be more robust against sample group differences than empirically derived reference distributions – if the theory is valid. The multinomial distribution, a form of discrete11 probability distribution, can be used to describe the distribution of all possible outcomes under the null hypothesis to the analytic theory of the CQT.

Analytic theory of the CQT

The analytic theory of the polygraph has been discussed and evaluated in numerous studies and publications (Bell, Raskin, Honts & Kircher, 1999; Honts & Peterson, 1997; Honts & Raskin, 1988; Honts & Reavy, 2015; Kircher & Raskin, 1988; Kircher, Packard, Bell & Bernhardt, 2001; MacLaren & Krapohl, 2003; Nelson, 2014, 2015a, 2015b; Raskin, Honts & Kircher, 2014; Raskin, Kircher, Honts & Horowitz, 1988), and holds that greater changes in physiological activity are loaded at different types of test stimuli as a function of deception or truth telling in response to the

relevant target stimuli. During the interview phase of a polygraph examination an examinee who does not wish to make a confession will deny involvement in a behavioral issue under investigation.

During a polygraph test, changes in physiological activity are recorded using an array of recording sensors. Data from the recording sensors is subject to numerical transformation and reduction for statistical analysis. The goal of the analysis is to classify test results as deceptive or truthful based on the differential salience (Handler & Nelson, 2007; Senter, Weatherman, Krapohl & Horvath, 2010) of different types of test stimuli. The psychological basis for observed differences in physiological activity can be thought of as generally involving a combination of the mental effort necessary to conceal the truth and assert a lie, emotion related to the behavioral act or the potential consequences for the act, and conditioned responding to the descriptive stimulus as a result of involvement or experience in a behavioral act (Hander, Shaw & Gougler, 2010; Nelson, 2015a) under investigation.

Polygraph testing is neither a deterministic observation of deception or truth telling (i.e., perfect or unchangeable and not amenable to human behavior), nor a direct physical or linear measurement of deception or truth. Scientific tests are not expected to be infallible and are fundamentally probabilistic – including when probabilistic results are reduced to categorical results for convenience. Like other scientific tests, the purpose of the polygraph test is to record and analyze data as a basis for replicable calculation of the probabilistic result (American Polygraph Association, 2011, Nelson & Handler, 2012, 2015; Nelson, et al., 2011). Probabilities associated with test results can refer to expected classification accuracy rates with groups or samples of exams, and can also refer to the estimated statistical error or accuracy level for a single examination.

---

7 A statistical distribution is a set of numbers that can represent a phenomenon of interest (e.g., height, weight or polygraph scores) for which the data are nondeterministic or imperfect and are expected to vary somewhat. Data that vary in a completely unordered or random manner will not be useful to guide our conclusions about observations of real world phenomena. Data that vary with some degree of order can be useful if the rules and assumption that determine the form of the data distribution can be studied and proofed by statisticians and mathematicians. Statistical distributions are characterized by numerical parameters that provide all the information necessary to calculate the distribution mathematically.

8 For example: the Gaussian or normal distribution, sometimes called a bell curve, is a commonly used theoretical distribution that is related to the standard normal or z distribution. The normal distribution characterizes a variety of naturally occurring phenomena. There are a number of other common and recognizable theoretical distributions, including the Chi squared distribution that is the sum of squared standard normal deviates, the t distribution that characterizes the distribution of small samples and which will converge towards the normal distribution for large samples, the binomial or Bernoulli distribution for discrete values that will be asymptotically normal for large sample sizes, the Poisson distribution that characterize the frequency of occurrence of time series events, the Weibull distribution that can be used to characterize the reliability of lifetime and failure events in engineering, the family of exponential logarithmic distributions that can be used to characterize nonlinear increases or decreases in events, the uniform distribution of decimal proportions between 0 and 1, and other theoretical distributions.

9 For example: sampling data that are normally distributed will produce a histogram of similar shape to the standard normal distribution. However, whereas a histogram is a description of available empirical sampling data, a theoretical distribution such as the standard normal distribution is a mathematical abstraction.

10 Statistical procedures often involve the study of an observed empirical distribution with reference to a theoretical statistical distribution that is a mathematical abstraction. When the empirical data conform reasonably to the shape of a theoretical distribution we can then use our mathematical knowledge of the theoretical distribution as a model to make replicable probabilistic and categorical inferences about our empirical data. When the empirical data are randomly selected or representative of the population from which the data was drawn we can begin to make inferences about the population from which the empirical sample was obtained.

11 A distribution is said to be discrete when the numerical values cannot be divided into fractions or smaller parts, when there are no meaningful values in between the nodal values that are characteristic of the data. For example: a person's height or weight can be expressed in continuous numerical values including decimals or fractions, while the number of times a person gets kicked by a horse can be expressed using only positive integers for which there is no meaningful interpretation in between each integer. Theoretical distributions are said to be continuous when the data values can be expressed using numbers than can be continuously divided into infinitely smaller and smaller parts for which there remains some useful and meaningful interpretation. For example, the uniform distribution of probabilities between 0 and 1 is a continuous distribution.

Administration and scoring of the CQT

The CQT is administered through the use of a no accusatory pretest interview, during which the issue under investigation is clarified and all test questions are reviewed with the examinee (American Polygraph Association, 2016; Raskin & Honts, 2002; Raskin, Honts & Kircher, 2014, Handler & Nelson, 2008), followed by the acquisition and recording of the test data in response to several iterations of a sequence of stimulus questions that includes the relevant or investigation target stimuli, comparison question stimuli (Kircher and Raskin, 1988; Bell et al., 1999; Department of Defense, 2006 Handler & Nelson, 2008; Krapohl & Shaw, 2015) and other procedural questions. A common CQT question sequence for an event specific diagnostic exam will include three relevant target questions, and three comparison questions, and will be repeated three to five times (Bell, Raskin, Honts & Kircher, 1999; Department of Defense, 2006, Handler & Nelson, 2008; Krapohl & Shaw, 2015). CQT data consist traditionally of timeseries recordings from three different sensors, including the thoracic and abdominal respiration sensors, an electrodermal activity (EDA) sensor and a cardiovascular activity sensor. A vasomotor sensor, also sometimes referred to as a photoelectricplethys mograph (PLE or PPG), can also be included. Data are transformed to numerical scores for each stimulus presentation and each recording sensor.

Physiological responses to CQT stimuli are coded using a nonparametric rubric. By convention, positive scores are assigned to CQT responses when there is a greater change in physiological activity in response to the comparison stimuli, while negative scores are assigned when there is a greater change in physiology in response to the target stimuli. Scores of zero can occur when there is no interpretable difference in response, or when there is no response to both relevant and comparison stimuli, or when the data are not interpretable due to physical activity or other data artifact (Department of Defense, 2006; Krapohl & Shaw, 2015; Nelson, Krapohl & Handler, 2008; Nelson et al., 2011). The number of scores will be determined by the number of relevant questions, the number of sensors and the number of repetitions of the question sequence.

When using the ESS, EDA scores are weighted more than the other sensor scores. This is because EDA data has been shown to be more strongly correlated with differences between deceptive and truthful examinees and contributes more information to an optimal test model than other sensor data (Ansley & Krapohl, 1999; Honts, Handler, Shaw & Gougler, 2015; Harris, Horner & McQuarrie, 2000; Kircher, Kristjansson, Gardner & Webb, 2005; Kircher and Raskin, 1988; Krapohl & McManus, 1999; Nelson, Krapohl & Handler, 2008; Podlesny & Raskin, 1978; Podlesny & Truslow, 1993; Raskin, Kircher, Honts & Horowitz, 1988). The procedure for weighting the EDA scores is simply to double all EDA integer score values. EDA scores are therefore 2, 0, and + 2 when using the ESS, whereas scores from the other sensors are 1, 0 and +1. In this way, nonparametric ESS scores are intended to approximate an optimal statistical function. This is different than other manual scoring methods for which the data from various sensors are assumed to contribute equally to the effectiveness of the classification model.

Calculation of the multinomial reference distribution

Computation of the theoretical distribution of ESS scores begins with a statement of the null hypothesis to the analytic theory of the CQT. The null hypothesis says that physiological responses are not systematically loaded for target or comparison stimuli, and instead occur in a random manner for each of the recording sensors. Both the analytic theory and the null hypothesis pertain to the data and distribution of scores for the individual sensors in the same manner that these pertain to the grand total and question subtotal scores. It is expected that random data, under the null hypothesis, will give results that are meaningless and unpredictable, and this will be observed in classification accuracy rates that will not differ from random chance. The theoretical distribution of ESS scores is multinomial because there are more than two possible scores for each sensor at stimulus presentation (referred to more generally as a stimulus trial): 1, 0, and +1. Under the null hypothesis the sensor scores are not loaded in any systematic way, and are therefore uniformly or equally likely to occur 12.

For each recording sensor, there will exist a multinomial distribution of possible sensor totals determined by the number of trials and the number of possible sensor scores for each stimulus trial. For example: the sensor distributions for an event specific polygraph test with three relevant questions and three repetitions of the question sequence will consist of nine stimulus trials for each question for which there will be three possible sensor scores at each trial (27 sensor scores). Similarly, the sensor distribution for an event specific polygraph examination with three repetitions of a question sequence that includes only two relevant questions will consist of six stimulus trials, again with three possible sensor scores at each trial (18 sensor scores). In the same way, the sensor distributions for an event specific polygraph test with five repetitions of a question sequence that includes four relevant questions will consist of 20 stimulus trials with three possible sensor scores at each trial (60 sensor scores).

Some polygraph examinations are evaluated using only the question subtotal scores; in this case the number of stimulus trials for the calculation of the multinomial sensor distribution will be determined by the number of repetitions of the test question sequence. For example: the multinomial distribution for the sensor subtotal scores of a multiple issue polygraph with three repetitions of the question sequence will be calculated from three stimulus trials regardless of the number of relevant stimuli.

---

12 The multinomial distribution can also be calculated with weighted probability values for the possible sensor scores for each stimulus trial when there is a satisfactory basis of information to inform those probability values.

Computation of the multinomial distribution for sensor totals

A complete discussion of multinomial calculations is beyond the scope of this paper.

However, a worked example can be useful to illustrate the basic idea. Because transformed numerical results for all sensors can receive one of three possible values for each stimulus trial, all sensor distributions are identical under the null hypothesis. First it is necessary to establish a coherent vocabulary to describe the various ways of summarizing the numerical scores. Table 1 shows a sample score sheet, with simulated random data, illustrating the calculation of the question subtotals, sensor subtotals, sensor totals, and grand total score13. There are nine different multinomial distributions that can be calculated for the sensor total scores depending on the CQT format. This is because CQT formats can consist of two, three, or four relevant questions, and can be completed with three, four, or five repetitions of the test question sequence. A multinomial sensor distributions can be calculated for the sensor subtotals, for use when polygraph decision rules make use of question subtotals.

The score sheet in Table 1 shows an exam with 9 stimulus trials (i.e., there are three repetitions of a question sequence that includes three relevant questions). There are 19,683 unique permutations14 of the score sheet in Table 1 and 55 unique unordered combinations15 of the number of +, , and 0 scores. The number of unique permutation is calculated as n raised to the k power (n^k) where n is the number of different possible scores and k is the number of trials. Permutations are unique ordered sequences, and are not the same as combinations. The number of unordered combinations is calculated as (n+k1)! / (k! * (n1)!) where the "!" indicates the factorial16. The number of possible sensor scores for each multinomial sensor distribution is a function of the number of stimulus trials using this formula: 2*k+1, where k is the number of stimulus trials. For example, the multinomial distribution for the sensor totals with nine trials will include 19 possible values (2*9+1=19) for the sensor totals, ranging from 9 to +9 including the value 0.

---

13 The term sensor subtotal refers to the sum of the repetitions of each individual relevant question for a recording sensor. Sensor total refers to the sum of all scores for all repetitions of all relevant questions for a recording sensor. The sum of the sensor subtotals will equal the sensor total. The term grand total is used to refer to the sum of all sensor scores for all repetitions of all relevant questions. Question subtotal refers to the sum of all sensor scores for all repetitions of each individual relevant question. The sum of the question subtotal scores will equal the grand total score. There is no mathematical use for the subtotals for each presentation of each stimulus, nor for the sensor subtotals for each repetition of the stimulus question sequence in calculation of the multinomial distribution of CQT scores.

14 Permutations are unique ordered sequences of the 9 scores which consist of the values +, , and 0. Permutations are immutable, which means that the positions of the elements of a permutation are not interchangeable. In other words, the permutation (1, 2, 3) is not the same as the permutation (3, 2, 1) or (2, 1, 3) or any other order of the same values.

15 Combinations are sequences of items that are mutable, meaning that the positions of the items in the sequence can be moved without changing the value of the sequence. In other words, the combination (1, 2, 3) is the same as (3, 2, 1) because the order of the items is different though the items themselves are the same.

*16 The general form of the combinatoric formula is n! / ((nk)!*k!) for which common examples have k smaller than n. Factorial calculations can quickly become large and unwieldy making algebraic conventions useful. For example: how many unique groups of 3 persons can be made from 10 persons? Answer: 10!/(7!*3!) = (10 * 9 * 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1)*
*/ ((7 * 6 * 5 * 4 * 3 * 2 * 1) * (3 * 2 *1)) = (10 * 9 * 8) / (3 * 2 * 1) = 720 / 6 = 120. The number of k trials in the polygraph context is not constrained by and can exceed the value of n. For this reason, we use a different version of the formula.*

Table 1. Sample score sheet with question subtotals, grand total, sensor subtotals, and sensor totals.

Repetition 1	R1	R2	R3
Respiration	1	0	1
EDA	0	2	2
Cardio	1	0	0
Vasomotor	0	1	0
Repetition 2	R1	R2	R3
Respiration	0	1	0
EDA	2	0	2
Cardio	1	0	1
Vasomotor	1	0	1
Repetition 3	R1	R2	R3
Respiration	1	1	1
EDA	0	2	2
Cardio	0	1	1
Vasomotor	0	1	0
Question subtotals	3	1	1
Sensor subtotals	R1	R2	R3
Respiration	0	2	0
EDA	2	0	2
Cardio	2	1	2
Vasomotor	1	0	1
Sensors	Sensor totals	blank	
Respiration	2		
EDA	4		
Cardio	1		
Vasomotor	0		
Grand total	5		

The number of permutations and combination differs greatly for different CQT formats. Table 2 shows the number of unique permutations, unordered combinations and the number of different possible sensor scores for different CQT formation. The number of unique permutations can be thought of as the total number of different arrangements of scores that could possible occur on the score sheet as shown in Table. 1. The number of permutations can become quite large. For example: the sensor totals for a polygraph test with 20 stimulus trials (i.e., five repetitions of a question sequence that includes four relevant questions) will include 51 possible values (2*20+1=41) for which there are 3,486,784,401 unique permutations and 231 unordered combinations. In contrast, the sensor totals for a polygraph test with 6 stimulus trials (i.e., 3 repetitions with 2 relevant questions) will included 13 possible values from 6 to +6 for which there are 729 possible permutation with 28 unordered combinations of the number of +, and 0 scores.

The sensor total for a single relevant question in a multiple issue exam consisting of three repetitions of the test question sequence will include seven possible values (2*3+1=7) for which there are 27 unique permutations and 10 unordered combinations.

Because combinations are unordered (i.e., the location of the scores in the score sheet is mutable or changeable), the combinations of the number of +, and 0 scores, and resulting sensor totals, will occur more frequently than others. Returning to the example of a CQT with three repetition of a question sequence that includes three relevant questions, with the 19,683 unique permutations of the possible scores +, and 0, there is only one way to achieve a particular sensor score of +9 because all three repetitions of all three relevant questions must produce a sensor score of +1 to achieve this sensor score. Similarly, there is only one way to achieve a sensor score of 9. However, there are 3,139 different ways to achieve a sensor score of 0.

Table 2. Unique permutations (unordered combinations) and [different number of scores] for sensor totals.

	2 RQs	3 RQs	4 RQs	1 RQ (subtotal)
3 repetitions	729 (28) [13]	19,683 (55) [19]	531,441 (91) [25]	27 (10) [7]
4 repetitions	6561 (45) [17]	531,441 (91) [25]	43,046,721 (153) [33]	81 (15) [9]
5 repetitions	59,049 (66) [21]	14,348,907 (136) [31]	3,486,784,401 (231) [41]	243 (21) [11]

Calculation of the multinomial distributions for CQT scores requires the enumeration of all possible permutations and combinations. With very small data sets the permutations can be enumerated manually – sometimes even mentally when the data are very tiny. The advantages of larger data sets are several, and include smaller errors of measurement, greater precision, and reduced granularity of the numerical results. It will be simpler and more expedient to work with combinations, instead of permutations, whenever possible when the datasets become larger. This is the purpose of combinatorics and multinomial calculations.

To calculate multinomial reference table for sensor scores all that is necessary is to know the number of possible scores for each trial (+1, 1, 0), the probability weights associated with each possible score (.333, .333, .333)17, and the number of k trials that will be used (number of relevant questions * number of repetitions). In this example, the number of relevant questions is three and the number of repetitions is also three, and so the number of k trials is 9. To calculate the number of ways to achieve each score it is first necessary to enumerate all 55 possible combinations of sensor scores (i.e., how many scores of +1, 1, and 0)18, and then sum the scores for each combination and calculate the factorial for the result. Next it will be necessary to calculate the factorial for the product of the scores for each combination. Finally, we can divide the factorials of the sums by the factorials of the products. The result will be the number of ways to achieve each combination of scores. Each of the 55 combinations of scores must be summed after multiplying the number of each possible score by the value of the score, and it will be noticed that the sums will be

similar for some combinations. By summing the number of ways for all similar sums we can determine the total number of ways to achieve each of the 19 possible sensor scores.

As stated earlier, there are 12 multinomial sensor distributions needed for the ESS, including 9 distributions for the sensor totals and 3 for the sensor subtotals. These distributions will describe the number of ways to achieve each of the possible sensor scores along with the proportion of ways to achieve each score compared to the distribution. Although mathematical concepts are themselves simple, the calculate of all the ways to achieve all the possible sensor totals and sensor subtotals for all polygraph test format could become a laborious and punishing task if one attempts to do this manually. Fortunately, programmable computers and statistical software are available today, and can reduce the arduousness of these calculations for us when we know the correct formula and procedure.

The probability mass function19(pmf) can be calculated by taking the total number number of ways to achieve each possible sensor score and dividing that by the total number of different possible sensor scores. The pmf of each sensor score will be used later as the probability weight for the possible sensor scores when calculating the multinomial distribution of the combined sensor scores. Table 3 shows the multinomial sensor table for a polygraph test with three repetitions of three relevant questions, including the number of ways to achieve each possible sensor score and the probability mass function for each score.

---

17 These probabilities are uniform because the multinomial distribution of sensor scores is calculated under the null hypothesis that greater changes in physiology are not systematically loaded and are instead randomly distributed, resulting in uniform probabilities for their occurrence.

18 For example: if there are nine scores of +1 then there can be zero scores of 1 or 0. If there are eight scores of +1 then there can be one score of 1 and zero scores of 0, or one score of 0 and zero scores of 1. And so on.

19 The probability mass function describes the proportion of scores at each level in the distribution and can be used to estimate the likelihood of achieving a particular score under the null hypothesis.

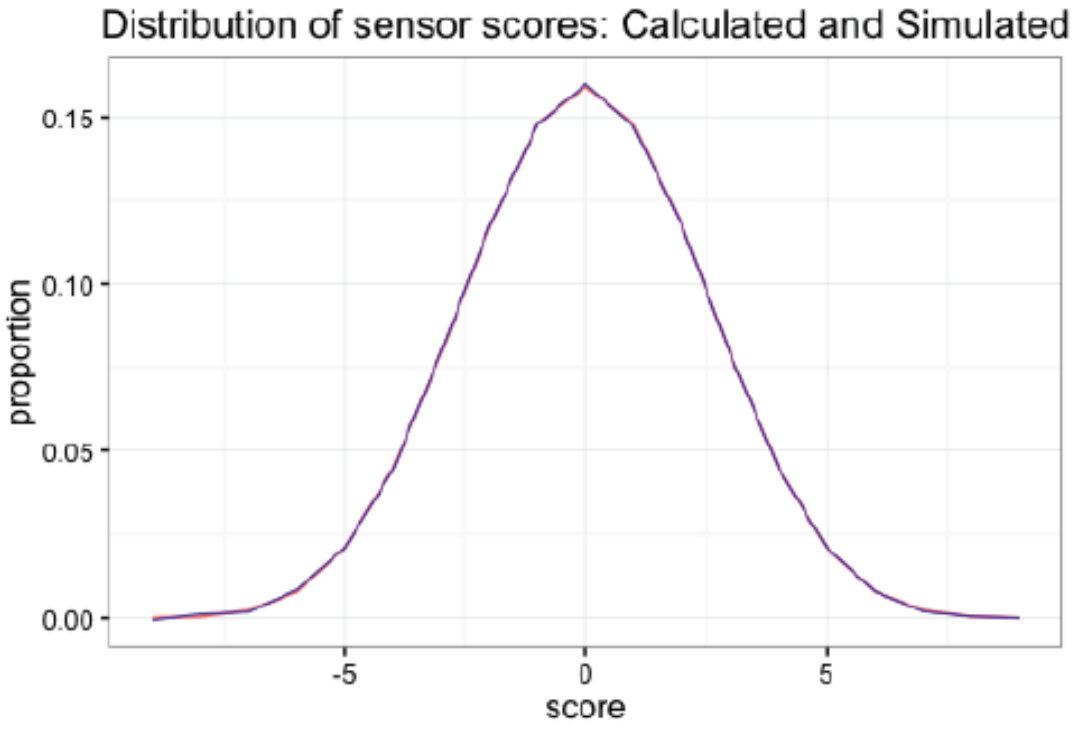
20 The probability mass function describes the proportion of scores at each level in the distribution and can be used to estimate the likelihood of achieving a particular score under the null hypothesis.

Table 3. Multinomial for one sensor total with three repetitions of three relevant questions.

score	ways	*Pmf*
-9	1	.0001
-8	9	.0005
-7	45	.0023
-6	156	.0079
-5	414	.0210
-4	882	.0448
-3	1554	.0790
-2	2304	.1171
-1	2907	.1477
0	3139	.1595
1	2907	.1477
2	2304	.1171
3	1554	.0790
4	882	.0448
5	414	.0210
6	156	.0079
7	45	.0023
8	9	.0005
9	1	.0001

The pmf in Table 3 was compared to a simple Monte Carlo simulation of 1 million iterations of a sample space consisting of n=9 random selections from the uniform distribution of [.333, .333, .333]. Each set or iteration of nine random selections resulted in a sum between 9 and +9, for which the results were aggregated over the 1 million iterations to determine he proportion of results that produced each of the possible integer scores between 9 and +9. Results of the comparison between the closed form multinomial calculation of this distribution and the Monte Carlo simulation are shown in Figure 1. There is virtually perfect concordance between the distributions, and differences are made visible only through the addition of a small amount of noise to one of the lines. The meaning of this is that the multinomial calculations can be considered correct because they can be verified with a simulation for which the intuition is simpler than the intuition for the combinatoric math.

Figure 1. Histogram comparing a MonteCarlo simulation with the closed form multinomial sensor distribution for three repetitions of a sequence that includes three relevant questions.



Reference tables for sensor totals are shown in Appendices AC for CQT formats consisting of five repetitions of question sequences including two, three and four relevant questions with three position scoring. Appendix D shows the reference table for sensor scores for sensor subtotals with five repetitions of the relevant questions with three position scoring.

Computation of the multinomial reference distribution for combined sensor scores

Because no classification can be made using an individual sensor total, the distribution of combined sensor scores will be of more useful to field examiners than the distribution of scores for individual sensors. The distribution of combined CQT sensor scores is the combination of the multinomial distributions of the scores for the individual sensors using the pmf for the sensor scores as the weighting coefficients.

One important aspect of the multinomial distribution of ESS scores is that EDA scores are weighted more than other sensor scores in attempt to approximate a more optimal statistical function than can be achieved by naïve weighting21. A consequence of this weighting is that sensor totals for ESS scores are immutable (i.e., scores cannot

be interchanged for the sensors) and the multinomial distribution cannot be calculated using the computationally more convenient method involving unordered combinations of sensor scores. Instead the multinomial distribution of ESS scores must be completed using the more exhaustive method involving unique permutations. Fortunately, computers today, once they are given properly coded instructions, can perform this task easily.

---

21 Naive in this usage (analytics and statistics) refers to an assumption, not necessarily supported by evidence, that we know nothing about the relative importance and contribution of the different sensor data to the final test result and precision of the test model.

The distribution of the combined sensor totals – in the form of a grand total score or question subtotal score – is also a multinomial distribution. For the distribution of combined sensor scores n is the number of possible scores that can result for each sensor (e.g., n = 19 possible scores ranging from 9 to +9 when there are three repetitions of a question sequence that includes three relevant questions) while k is the number of sensors included in the grand total or question subtotal scores (i.e.., k = 4 when using the respiration, EDA, cardio and vasomotor sensors). So, the distribution of the combined sensor totals will be determined by the number of stimulus trials and the number of sensors. The likelihood of each is expressed by the pmf for the sensor total (as shown in Table 3).

The pmf for the sensor total therefore gives us the probability weights for the calculation of the multinomial distribution for combined sensor scores. The distribution of combined sensor scores will have a range of 2*n+1 where n is the product of the number of sensors and the number of stimulus trials. Following the same example that was started earlier, the grand total score for an exam with three repetitions of a question sequence that includes three relevant questions and four recording sensors (respiration, EDA, cardio, and vasomotor) will have a range of 91 possible ESS scores. This is because the four recording sensors have a combined maximum of score of 5 for each stimulus trial, because ESS EDA scores are weighted more than the other sensor scores, and because (5*9=45) while 2*45+1=91. This multinomial distribution has 130,321 unique permutations.

Table 4 shows the multinomial distribution of ESS grand total scores for a polygraph with three repetitions of three relevant questions using the traditional array of sensors (respiration, EDA, cardio), including the range of possible scores, number of ways to achieve each score and the pmf for each CQT score. Also shown in Table 4 is the cumulative distribution function22(cdf), continuity corrected pmf23, along with the odds24. Finally, because point estimation is realistically less useful than interval estimation, the 5th percentile lower limit of the confidence interval was calculated for the odds using the Clopper Pearson method25 for the binomial (Agresti & Coull, 1998; Clopper & Pearson, 1934; Newcombe, 1998; Thulin, 2014). The lower limit of the Clopper Pearson interval allows us to estimate the proportion of repeated experiments that can be expected to exceed a threshold if the present data are informing us correctly about reality. When used in the context of Bayesian decisionmaking, the Clopper Pearson interval may be thought of as a credible interval that describes the level of confidence or uncertainty about a probabilistic and categorical conclusion26,27.

---

22 The cumulative distribution function is the cumulative sum of the pmf.

23 The continuity correction is calculated by averaging all pairs of cell values. This has the effect of placing the location of the probability value in the middle of the cell instead of at the edges. This is analogous to sports betting wherein a bet is place on a point value such as 55.5 even though ½ points are never scored in reality. This allows a more straightforward discussion of the odds that the actual point score will be over or under the value.

24 Odds are always presented as relative to the value of 1 and indicate the likelihood of achieving a score of equal or more extreme value.

25 This interval estimation method was selected because it known to never have less than the nominal coverage area. In other words, the actual coverage rate for a 95% confidence interval may exceed 95% depending on the input parameters. Other interval estimation methods may have actual coverage rates that are less than nominal depending on the input.

24 Credible intervals in Bayesian analysis are analogous to confidence intervals in frequentist analysis, except that Bayesian analysis regards the criterion of interest as a probability and the data as fixed (i.e., it is the information available with which to calculate a conclusion). In contrast, frequentist confidence intervals regard the criterion as fixed (reality exist in only one form) and regards the data as a random variable for which the confidence interval describes the likelihood of obtaining the data.

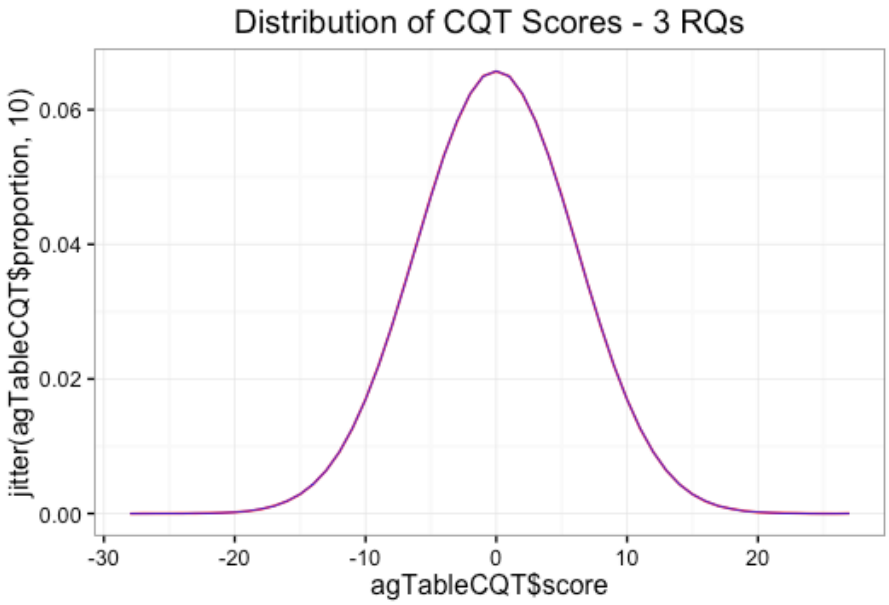
25 For example: the lower limit of a Bayesian credible interval might tell us that we are 95% certain that the odds exceed a particular value.

Table 4. Multinomial distribution of ESS grand total scores for three repetitions of a question sequenced that includes three relevant questions, with the number of ways to achieve each score, pmf, cdf, continuity-corrected cdf, odds and the 5th percentile lower limit of the Clopper-Pearson interval (extreme values are omitted).

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-19	90	.0004	.0008	.0006	1712	11.27
-18	100	.0007	.0014	.0011	910.8	11.07
-17	108	.0011	.0025	.0020	503.7	10.73
-16	117	.0018	.0043	.0035	288.9	10.21
-15	124	.0029	.0072	.0058	171.4	9.47
-14	132	.0044	.0116	.0094	105.1	8.52
-13	138	.0064	.0179	.0148	66.37	8.45
-12	145	.0092	.0270	.0227	43.11	7.08
-11	150	.0127	.0394	.0336	28.73	6.28
-10	156	.0169	.0558	.0485	19.62	5.31
-9	160	.0220	.0771	.0681	13.69	4.35
-8	165	.0278	.1037	.0931	9.74	3.6
-7	168	.0341	.1360	.1242	7.05	2.9
-6	172	.0406	.1742	.1617	5.18	2.34
-5	174	.0471	.2181	.2057	3.86	1.87
-4	177	.0531	.2673	.2558	2.91	1.48
-3	178	.0584	.3211	.3115	2.21	1.17
-2	180	.0624	.3786	.3717	1.69	0.91
-1	180	.0649	.4386	.4350	1.3	0.71
0	181	.0658	.5000	.5000	1	0.55
1	180	.0649	.5614	.5650	1.3	0.71
2	180	.0624	.6214	.6283	1.69	0.91
3	178	.0584	.6789	.6885	2.21	1.17
4	177	.0531	.7327	.7442	2.91	1.48
5	174	.0471	.7819	.7943	3.86	1.87
6	172	.0406	.8258	.8383	5.18	2.34
7	168	.0341	.8640	.8758	7.05	2.9
8	165	.0278	.8963	.9069	9.74	3.6
9	160	.0220	.9229	.9319	13.69	4.35
10	156	.0169	.9442	.9515	19.62	5.31
11	150	.0127	.9607	.9664	28.73	6.28
12	145	.0092	.9731	.9773	43.11	7.08
13	138	.0064	.9821	.9852	66.37	8.45
14	132	.0044	.9885	.9906	105.1	8.52
15	124	.0029	.9928	.9942	171.4	9.47
16	117	.0018	.9957	.9966	288.9	10.21
17	108	.0011	.9975	.9980	503.7	10.73
18	100	.0007	.9986	.9989	910.8	11.07

The distribution shown in Table 4 was also compared a simple MonteCarlo simulation of 1 million iterations. The MonteCarlo simulation for ESS grand totals scores consisted of three sensor scores with a range of -9 to +9, which were sampled using the pmf from Table 3 as the sampling weighting coefficients. After multiplying the EDA scores by two, the sum for each case in the simulation was an integer between 36 to +36. Results for the 1 million simulations were aggregated for the number and proportion of iterations that produced each of the possible scores from -36 to +36. A comparison between the closed form multinomial calculation of this distribution and the Monte Carlo simulation is shown in Figure 2. There is again virtually perfect concordance between the distributions, and differences are made visible only through the addition of a small amount of noise to one of the lines.

Figure 2. Histogram comparing a Monte Carlo simulation of CQT scores with the closed form calculations of the distribution of ESS grand total scores three repetitions of a sequence that includes three relevant questions using the respiration, EDA, and cardio sensors.



Appendices IK show the multinomial reference distributions for grand total scores of CQT question sequences that include two, three, and four relevant questions with the addition of the vasomotor sensor. Appendix L shows the multinomial reference distribution of CQT subtotal scores using the addition al vasomotor sensor. These reference tables can serve as the likelihood function for naïveBayes classification methods, and may be of interest to those who wish the study or replicate the closed form multinomial calculation or to compare these results with simulation. Reference tables such at that shown in Table 4 and those in the appendices can be used to determine the cutpoints for statistical significance prior to testing, and can also be used to the statistical values associated with a test result.

Determination of the cutpoints using Table 4 is a matter of looking in the last column, for lower limit of the Clopper Pearson interval for the odds, and then selecting the smallest value that is greater than 1 along with the largest value that is less than 1 and then looking in the first column to determine the cutpoint for those odds. Table 4 shows that cutpoints of +3 and 3 exceed the odds 1 and 1, meaning that scores that equal or exceed these cutpoints are significantly likely to improve our knowledge if we begin by assuming we know nothing. Table 4 can also be used to determine the odds associated with a test score. To do this, simply locate the lest score in the leftmost column and

then select corresponding value from the last column for lower limit of the Clopper Pearson interval where scores that exceed the values 1 and -1 are statistically significant at the .05 (one tailed) level.

Reference tables such as that shown in Table 4 can reduce the need for procedurally intensive recalculation of a range of values that may be used repeatedly. For this reason, to reduce the computational workload for those who wish to study or work with the multinomial distributions for CQT scores, all ESS distributions of interest to polygraph formats in field practice use today can be calculated and saved in a series reference tables. Appendices EG show the multinomial reference distributions for grand total scores of CQT question sequences that include two, three, and four relevant questions using the traditional array of respiration, EDA and cardio sensors. Appendix H shows the distribution of CQT subtotal scores along with multiplicity corrections28 for two, three and four relevant questions29 scores using the traditional sensor array. Appendices IK show the multinomial reference distributions for grand total scores of CQT question sequences that include two, three, and four relevant questions with the addition of the vasomotor sensor. Appendix L shows the multinomial reference distribution of CQT subtotal scores with multiplicity correction for two, three and four relevant questions using the additional vasomotor sensor. These reference tables can serve as the likelihood function for naïveBayes classification methods, and may be of interest to those who wish the study or replicate the closed form multinomial calculation or to compare these results with simulation.

Comparison of Table 4, for three repetitions of a question sequence with three relevant questions, with the one in Appendix F, for five repetitions of a question sequence with three relevant questions, shows that although the statistical values may differ slightly the integer cut scores are identical. For this reason, the tables in Appendices EL show only the calculations with five repetitions. Table 5 shows the ESS cut scores for statistical significance for event specific polygraphs using the multinomial reference distributions with a one tailed alpha = .05 for the lower limit of the Clopper Pearson interval for both positive and negative classifications. Table 6 shows the ESS cut scores for subtotal scores of polygraphs interpreted with an assumption of independent criterion variance. Inspection of Tables 5 and 6 indicate that integer cut scores, determined by the lower limit of the Clopper Pearson interval, are different when using the vasomotor sensor.

Table 5. ESS cut scores for grant total scores of event specific exams using the multinomial reference distributions, using a one tailed alpha = .05 for the lower limit of the Clopper Pearson interval for positive and negative classifications (multiplicitycorrected subtotal cut scores in parenthesis).

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)
Respiration, EDA, Cardio, Vasomotor	+3 / -3 (-5)	+3 / -3 (-7)	+3 / -3 (-9)

28 The multiplicity corrected odds were calculated as the exponent of the natural log of the subtotal odds divided by the number of relevant questions raised the sign value of the lowest subtotal score [exp(log(minSubtotal odds)/numberRQs^minSubtotalSign].

29 In practice only the lowest subtotal score is used for classification though the multiplicity correction is calculated as a function of the number of relevant questions.

Table 5. ESS cut scores for subtotal scores of multiple issue exams using the multinomial reference distributions, using a one tailed alpha = .05 for the lower limit of the Clopper Pearson interval without statistical correction for positive classifications and with statistical correction for negative classifications.

	2 RQs	3 RQs	4RQs
Respiration, EDA, Cardio	+2 / -3	+1 / -3	+1 / -3
Respiration, EDA, Cardio, Vasomotor	+2 / -3	+1 / -3	+1 / -3

Discussion

Counting things is an ancient human activity – perhaps the second, or third, or at least possibly among the top five of the oldest professions. Human progress and scientific progress can, in many ways, be thought of as a function of the improvements in our ability to count and quantify things. Combinatorics and multinomial calculations is simply a way to count things for which different possible combinations can exist.

To appreciate the importance of both theoretical and empirical probability distributions it is useful to remember the difference between the two. Empirical distributions are based on the observation of outcomes in a dataset for a population, sample or individual. On the other hand, theoretical probability distributions are based on a mathematical function that defines the distribution of values that could possibly occur within our theoretical understanding of the data. It will also be useful to remember that probability, in general, refers to the measurement of uncertainty and the chance of a given event occurring.

An overarching goal of science is to learn the general facts and principles about how reality and the universe works. But the volume of phenomena and data in the universe is far too great to work with, and so science often requires that we attempt to learn from sampling data. Inferential statistics and probability theory are intended to help us to determine what can be reasonably said about reality and the universe based on our analysis of the available data. In the context of the polygraph or other scientific test, statistics and probability theory is intended to help us determine what can be said about the test subject. Statistics is simply the mathematical language of science, because the goal of quantification related directly to the goal of scientific knowledge.

Inferential statistics begins by observing empirical data to determine the distribution of observed values, and ends by making reference to a theoretical distribution. Theoretical distributions are the core of statistical decision making because they allow us to make replica table mathematical estimations about important phenomena for which we can obtain neither a physical measurement nor perfect deterministic observation. Use of the term theoretical should not be misunderstood as implying speculation or impracticality. Theoretical distributions sit at the core of inferential statistics because they allow us to make allow us to make rational and replicable estimates and predictions about any phenomena for which no deterministic solution or physical measurement can be achieved.

Slightly different interpretations may be suggested by the use of empirical and theoretical distributions. Use of empirical distributions in the original ESS involved a pragmatic assumption that the test result belonged to one of two groups if the test score satisfies a specified probability threshold that defines the boundary of statistical significance for the opposing group. There are two empirical distribution because we are seeking one of two classifications. Evaluation of CQT data using a theoretical distribution depends on a single distribution calculated under the null hypothesis that CQT data are no systemic or meaningless, occurring only randomly. Instead of comparing the test data to a statistical thresh old for the opposing classification, use of the theoretical distribution requires the comparison of the test data, and the hypothesis that the data are systematically loaded as a function of deception and truth telling, against the null hypothesis of random responses. An interpretation of statistical significance can be made when the test score satisfies a decision or classification boundary that can be specified in terms of a proportion or odds ratio that describes the loading of the numerical scores and physiological responses to test stimuli. Both empirical and theoretical distributions can be used in Bayesian classification and decision models.

Theoretical reference distributions for CQT scores may provide a in important and useful and generalizable probability estimate for ESS scores. Whereas empirical reference distributions depend heavily on the representativeness of a volume of available sampling data, theoretical distributions depend more directly on the validity of the operational or analytic theory – that data are loaded systematically as a function of deception or truth telling. The value of an analytic and operational theory for the comparison question test is that answers to questions about validity rely more on observations about real world test performance than upon understanding the exact psychological or physiological mechanism that explain why the test works – though questions about psychological and physiological constructs will remain important areas for scientific inquiry and research.

Theoretical distributions rely on the mathematical expression, and mathematical proof, of our understanding of reality, and can be compared with the practical observation of existing empirical data. Effectiveness of any interpretation of the practical categorical meaning of the theoretical probability outcomes of polygraph test results will rest on both the correctness of mathematical expressions, and the correctness of the theoretical assumption that responses to different types of test stimuli do, or do not, vary as a function of truth telling or deception to the target questions. That nonsystematic and meaningless data can be characterized by random numbers is well proven to the point where it is accepted as axiomatic.

Summary

This project involved the calculation of theoretical reference distributions for ESS scores of CQT formats that consist of up to five repetitions of a question series that can include two, three, or four relevant questions, in addition to the calculation of the reference distributions for subtotal scores. The theoretical distribution of ESS scores for CQT data will take the form of a discrete multinomial distribution determined by the number of relevant questions, the number of repetitions of the test stimuli, and the number of physiological recording sensors. In probability theory, multinomial distributions provide the probability of observing any particular combination of items for a set of possible outcomes that are repeated multiple times.

Computation of the multinomial theoretical distribution for CQT scores begins with the calculation of the multinomial distribution of scores for the individual physiological recording sensors. The multinomial sensor distribution is a function of the number of possible outcomes for each stimulus trials. The number of stimulus presentations for individual sensors is a function of the number of relevant questions and the number of repetitions. Field practices require the use of three to five repetitions of the test questions. CQT formats for event specific polygraphs can include two to four relevant target stimuli. Test formats that are interpreted with an assumption of independent criterion variance can also include two to four target stimuli. The distribution of CQT test scores is the multinomial distribution of the combined multinomial distributions for the array of recording sensors. Recording sensors traditionally include the respiration, EDA and cardio sensors and can also include a vasomotor sensor.

Two versions of the multinomial reference data were calculated, using the traditional array of respiration, EDA and cardio sensors, and also with the addition of a vasomotor sensor. This represents an important advancement to the polygraph test because previously published scoring algorithms and previously published empirical reference tables did not include vasomotor sensor data. The addition of new sensor data to existing testing and analysis methods is a nontrivial endeavor.

Closed form calculations of the multinomial reference distributions were compared graphically with the results of Monte Carlo simulation, and showed the two methods can be expected to produce virtually identical distributional results. A general description of the calculation of the multinomial reference distributions is provided for replication and for readers who wish to develop their intuition and understanding of multinomial calculations and multinomial distributions.

Theoretical distributions can be useful to make replicable frequentist inferences about empirical data, and can also be useful as a likelihood function for Bayesian analysis. Whereas empirical distributions provide a basis for probabilistic estimation that an observed test data would be produce by a member of the population represented by an empirical reference distribution, theoretical distributions can provide a basis for a likelihood function in Bayesian analysis. Bayes analysis30 permits the inference of the cause of the data – which a more direct and intuitive conclusion about the probability that a polygraph test result was produced by a deceptive or truthful person.

The purpose of any scoring system is twofold. First it should attempt to optimize the effectiveness of the classification model and interpretation of test results31. Secondly, it should help to enable the computation of reasonable estimates of the probability that the classification is correct or incorrect. It is expected that any valid scoring or analytic method is supported by theoretical assumptions that can be clearly stated and expressed mathematically.

A scientific theory is an expression of our assumptions or conclusion of the universe, or some aspect of it, and tells us which aspects of our observation of the universe can be understood in a manner that is consistent with our understanding of other observations and other assumptions. The mathematical representation of a theory allows us to more reliably predict the consequences or results that can be expected to follow from the theory's assumption. An invalid theory, or rather an invalid hypothesis, will be useless. No amount of pretending will make an invalid hypothesis useful, and the only way to retain an invalid theory will be to disconnect from reality and engage or intellect in the practice of pseudoscience. If the analytic theory of the polygraph is correct, then a computational and intuitive understand of these multinomial reference distributions may be of some usefulness to both scientists and field practitioners.

An advantage of the theoretical distribution and a Bayesian approach is that the replacement or addition of evaluation features and recording sensors can be a simple matter when naïve assumptions are made. Use of theoretical distribution may also offer potential advantages such as robustness against group difference, and a simpler route towards the study and understanding of the empirical and practical value of the polygraph test result. Increasing the awareness and competence of polygraph professionals in the theory and application of theoretical reference distributions may lead to improved general understanding of the scientific meaning of polygraph test results, and may help to prevent incorrect interpretations and unrealistic expectations for deterministic perfection from probabilistic test results.

Availability of a theoretical distribution for ESS scores may help to advance the practical and empirical validity of the polygraph test by relieving concerns about the representativeness of available sampling data. This is because, unlike empirical distributions, theoretical distributions are mathematical abstractions that can be robust against some group differences as long as the basic analytic theory remains valid for different groups.

---

30 Bayesian analysis requires three elements: some data, a prior probability, and a likelihood function to apply to the test data in order to update the prior probability to a posterior probability. Prior probabilities are an important aspect of Bayesian analysis, but are not addressed in this manuscript.

31 Tests can be optimized for a number of purposes, according to operational priorities and mission objectives, including: test sensitivity, test specificity, falsepositive errors, falsenegative errors, positive predictive value, negative predictive value, or any other metric for test precision.

Finally, this project does not include an analysis of empirical data. It is limited to the mathematical calculation and simulation of the theoretical distributions of CQT scores under the null hypothesis to the operational or analytic theory of the polygraph test. Empirical evidence will still be required to demonstrate that classification into the criterion categories of guilt or innocence corresponds in the expected ways with differences in response to different types of test stimuli. Ultimately, the effectiveness of a classification method will always remain an empirical concern especially when the results may play a role in human decision making. It is hoped that the publication of this

description of the multinomial reference distributions, and corresponding reference tables for CQT scores, will help to advance the polygraph profession through the development of more objective, accountable and replicable analysis models. Of course, effective field polygraph examination may still continue to be subject to constraints and requirements around the test administration. And, as always, addition research is recommended.

References

Abramowitz, M. & Stegun, I. A. (1972). Multinomial Coefficients. §24.1.2 [pp. 823-824] in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 10th printing. National Bureau of Standards (now National Institute of Standards and Technology).

Agresti, A. & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. American Statistician, 52(2), 119-126.

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph, 40(4), 196-305. [Electronic version] Retrieved August 20, 2012, from http://www.polygraph.org/section/research-standards-apa-publications.

American Polygraph Association (2016). Standards of Practice. Approved by the Board of Directors on September 1, 2015. [Electronic version] Retrieved August 12, 2017, from http://www.polygraph.org.

Barland, G. H. (1985). A method of estimating the accuracy of individual control question polygraph tests. Proceedings of IDENTA, 1985, 142-147.

Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J.C. (1999). The Utah numerical scoring system. Polygraph, 28(1), 1-9.

Chen, C. C. & Koh, K. M. (1992). Principles and Technique in Combinatorics. World Scientific.

Clopper, C.; Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 26, 404–413.

Department of Defense (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retreived from http://www.antipolygraph.org/documents/federal-polygraph- handbook-02-10-2006.pdf on 3-31-2007. Reprinted in Polygraph, 40(1), 2-66.

Evans, M., Hastings, N. & Peacock, B. (2000). Statistical Distributions. 3rd ed. New York: Wiley. Handler, M. & Nelson, R. (2007). Polygraph terms for the 21st Century. Polygraph, 36, 157-164. Handler, M. & Nelson, R. (2008). Utah approach to comparison question polygraph testing. European Polygraph, 2, 83-110.

Handler, M., Shaw, P. & Gougler, M. (2010). Some thoughts about feelings: a study of the role of cognition and emotion in polygraph testing. Polygraph, 39, 139-154.

Harris, J., Horner, A. & McQuarrie, D. (2000). An evaluation of the criteria taught by the department of defense polygraph institute for interpreting polygraph examinations. Johns Hopkins University, Applied Physics Laboratory.

Honts, C., Handler M., Shaw P., & Gougler M. (2015). The vasomotor response in the comparison question test. Polygraph. 44(1), 62-78.

Honts, C. R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer. Available from the author.

Honts, C. & Raskin, D. (1988). A field study of the validity of the directed lie control question. Journal of Police Science and Administration, 16(1), 56-61.

Honts, C. R. & Reavy, R. (2015). The comparison question polygraph test: A contrast of methods and scoring. Physiology and Behavior 143, 15-26.

Kircher, J. C., Kristjiansson, S. D., Gardner, M. K. & Webb, A. (2005). Human and computer decision- making in the psychophysiological detection of deception. University of Utah, Salt Lake City.

Kircher, J. C. & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73, 291-302.

Kircher, J., Packard, T., Bell, B. G., & Bernhardt P. C. (2010). Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests. Polygraph. 39(1), 22-66.

Krapohl, D. J. (2002). Short report: Update for the objective scoring system. Polygraph, 31, 298-302. Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. Polygraph, 28, 209-222.

Krapohl, D. J. & Shaw, P, K. (2015). Fundamentals of Polygraph Practice. Elsevier.

MacLaren, V. & Krapohl, D. (2003). Objective assessment of comparison question polygraphy. Polygraph, 32, 107-126.

Nelson, R. (2014). What does the polygraph measure? APA Magazine 47(2), 39-47. Nelson, R. (2015a). Scientific basis for polygraph testing. Polygraph 41(1), 21-61.

Nelson, R. (2015b). One word essay (more or less): What does the polygraph measure? APA Magazine 48(2), 29-31.

Nelson, R. & Handler, M. (2012). Using normative reference data with diagnostic exams and the Empirical Scoring System. APA Magazine, 45(3), 61-69.

Nelson, R. & Handler, M. (2015). Statistical reference tables for comparison question polygraphs. Polygraph, 44(1), 91-114.

Nelson, R., Handler, M., Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B. & Oelrich, M. (2011). Using the Empirical Scoring System. Polygraph, 40, 67-78.

Nelson, R., Krapohl, D. & Handler, M. (2008). Brute force comparison: A Monte Carlo study of the Objective Scoring System version 3 (OSS-3) and human polygraph scorers. Polygraph, 37, 185-215.

Newcombe, R. G. (1998). Two sided confidence intervals for the single proportion: comparison of seven methods. Statistics in medicine, 17(8), 857-872.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., Clark, C. W. (2010). NIST Handbook of Mathematical Functions. National Institute of Standards and Technology (NIST), U.S. Department of Commerce, Cambridge University Press.

Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology, 15, 344-359.

Podlesny, J. A. & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. Journal of Applied Psychology, 78, 788-797.

Raskin, D. C. & Honts, C. R. (2002). The comparison question test. In M. Kleiner (Ed.), Handbook of Polygraph Testing. Academic Press.

Raskin, D. C., Honts, C. R. & Kircher, J. (2014). Credibility Assessment: Scientific Research and Applications. Elsevier.

Raskin, D. C., Kircher, J. C., Honts, C. R. & Horowitz, S.W. (1988). A study of the validity of polygraph examinations in criminal investigations. Final Report. National Institute of Justice, Grant No. 85-IJ-CX-0040.

Senter, S. M. & Dollins, A.B. (2003). New Decision Rule Development: Exploration of a two-stage approach. Report number DoDPI00-R-0001. Department of Defense Polygraph Institute Research Division, Fort Jackson, SC. Reprinted in Polygraph, 37(2), 149-164.

Senter, S., Weatherman, D., Krapohl, D. & Horvath, F. (2010). Psychological set or differential salience: a proposal for reconciling theory and terminology in polygraph testing. Polygraph, 39(2), 109-117.

Skiena, S. (1990). Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica. Addison-Wesley, Reading, MA.

Spiegel, M. R. (1992). Theory and Problems of Probability and Statistics. McGraw-Hill: New York.

Thulin, M. (2014). The cost of using exact confidence intervals for a binomial proportion. Electronic Journal of Statistics, 8, 817-840.

Wickens, T. D. (1991). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. Journal of Mathematical Psychology, 36, 213-234.

Wickens, T. D. (2002). Elementary Signal Detection Theory. Oxford.

Appendix A.

Multinomial Reference Distribution for Sensor Totals with 5 Repetitions of 2 Relevant Questions

score	ways	pmf
-10	1	<.0001
-9	10	.0002
-8	55	.0009
-7	210	.0036
-6	615	.0104
-5	1452	.0246
-4	2850	.0483
-3	4740	.0803
-2	6765	.1146
-1	8350	.1414
0	8953	.1516
1	8350	.1414
2	6765	.1146
3	4740	.0803
4	2850	.0483
5	1452	.0246
6	615	.0104
7	210	.0036
8	55	.0009
9	10	.0002
10	1	<.0001

Appendix B.

Multinomial Reference Distribution for Sensor Totals with 5 Repetitions of 3 Relevant Questions

score	ways	pmf
-15	1	<.0001
-14	15	<.0001
-13	120	<.0001
-12	665	<.0001
-11	2835	.0002
-10	9828	.0007
-9	28665	.0020
-8	71955	.0050
-7	157950	.0110
-6	306735	.0214
-5	531531	.0370
-4	827190	.0576
-3	1161615	.0810
-2	1477035	.1029
-1	1704510	.1188
0	1787607	.1246
1	1704510	.1188
2	1477035	.1029
3	1161615	.0810
4	827190	.0576
5	531531	.0370
6	306735	.0214
7	157950	.0110
8	71955	.0050
9	28665	.0020
10	9828	.0007
11	2835	.0002
12	665	<.0001
13	120	<.0001
14	15	<.0001
15	1	<.0001

Appendix C.

Multinomial Reference Distribution for Sensor Totals with 5 Repetitions of 4 Relevant Questions

score	ways	*pmf*
-20	1	<.0001
-19	20	<.0001
-18	210	<.0001
-17	1520	<.0001
-16	8455	<.0001
-15	38304	<.0001
-14	146490	<.0001
-13	484500	.0001
-12	1409895	.0004
-11	3656360	.0010
-10	8533660	.0024
-9	18062160	.0052
-8	34880770	.0100
-7	61757600	.0177
-6	100640340	.0289
-5	151419816	.0434
-4	210859245	.0605
-3	272290140	.0781
-2	326527350	.0936
-1	363985680	.1044
0	377379369	.1082
1	363985680	.1044
2	326527350	.0936
3	272290140	.0781
4	210859245	.0605
5	151419816	.0434
6	100640340	.0289
7	61757600	.0177
8	34880770	.0100
9	18062160	.0052
10	8533660	.0024
11	3656360	.0010
12	1409895	.0004
13	484500	.0001
14	146490	<.0001
15	38304	<.0001
16	8455	<.0001
17	1520	<.0001
18	210	<.0001
19	20	<.0001
20	1	<.0001

Appendix D.

Multinomial Reference Distribution for Sensor Subtotals with 5 Repetitions of the Question Sequence

score	ways	pmf
-5	1	.0041
-4	5	.0206
-3	15	.0617
-2	30	.1235
-1	45	.1852
0	51	.2099
1	45	.1852
2	30	.1235
3	15	.0617
4	5	.0206
5	1	.0041

Appendix E.

Multinomial Reference Distribution of ESS Grand Totals with 5 Repetitions of 2 Relevant Questions

score	ways	pmf	cdf	cdfContCor	odds	oddsLL05
-19	130	.0006*	.0014	.0011	903.2	12.29
-18	140	.0010	.0024	.0019	517.3	11.91
-17	148	.0016	.0041	.0033	305.8	11.33
-16	157	.0025	.0066	.0053	186.2	10.53
-15	164	.0038	.0103	.0085	116.7	9.51
-14	172	.0055	.0158	.0131	75.11	9.51
-13	178	.0077	.0234	.0198	49.57	8.04
-12	185	.0106	.0339	.0290	33.49	7.22
-11	190	.0142	.0477	.0415	23.13	6.18
-10	196	.0184	.0656	.0578	16.3	5.12
-9	200	.0233	.0880	.0788	11.69	4.29
-8	205	.0287	.1155	.1049	8.53	3.5
-7	208	.0345	.1482	.1367	6.32	2.82
-6	212	.0404	.1862	.1743	4.74	2.28
-5	214	.0462	.2293	.2177	3.6	1.84
-4	217	.0515	.2771	.2665	2.75	1.47
-3	218	.0561	.3290	.3202	2.12	1.16
-2	220	.0595	.3842	.3778	1.65	0.92
-1	220	.0617	.4415	.4382	1.28	0.72
0	221	.0625	.5000	.5000	1	0.57
1	220	.0617	.5585	.5618	1.28	0.72
2	220	.0595	.6158	.6222	1.65	0.92
3	218	.0561	.6710	.6798	2.12	1.16
4	217	.0515	.7229	.7335	2.75	1.47
5	214	.0462	.7707	.7823	3.6	1.84
6	212	.0404	.8138	.8257	4.74	2.28
7	208	.0345	.8518	.8633	6.32	2.82
8	205	.0287	.8845	.8951	8.53	3.5
9	200	.0233	.9120	.9212	11.69	4.29
10	196	.0184	.9344	.9422	16.3	5.12
11	190	.0142	.9523	.9586	23.13	6.18
12	185	.0106	.9661	.9710	33.49	7.22
13	178	.0077	.9766	.9802	49.57	8.04
14	172	.0055	.9842	.9869	75.11	9.51
15	164	.0038	.9897	.9915	116.7	9.51
16	157	.0025	.9934	.9947	186.2	10.53
17	148	.0016	.9959	.9967	305.8	11.33
18	140	.0010	.9976	.9981	517.3	11.91
19	130	.0006*	.9986	.9989	903.2	12.29
* extreme values ommitted						

Appendix F.

Multinomial Reference Distribution of ESS Grand Totals with 5 Repetitions of 3 Relevant Questions

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-22	360	.0009*	.0025	.0021	483	17.34
-21	370	.0013	.0038	.0031	317.7	16.38
-20	381	.0018	.0056	.0047	212.8	15.18
-18	400	.0035	.0115	.0098	100.6	13.93
-17	408	.0047	.0162	.0139	70.88	12.03
-16	417	.0062	.0223	.0193	50.72	11.12
-15	424	.0080	.0301	.0264	36.84	9.86
-14	432	.0102	.0402	.0355	27.14	8.48
-13	438	.0128	.0526	.0471	20.25	7.15
-12	445	.0157	.0680	.0613	15.31	6.13
-11	450	.0190	.0864	.0787	11.7	5.15
-10	456	.0226	.1081	.0996	9.04	4.27
-9	460	.0264	.1335	.1242	7.05	3.57
-8	465	.0304	.1624	.1526	5.55	2.99
-7	468	.0343	.1950	.1850	4.4	2.48
-6	472	.0382	.2310	.2213	3.52	2.05
-5	474	.0418	.2703	.2613	2.83	1.69
-4	477	.0449	.3125	.3046	2.28	1.4
-3	478	.0476	.3571	.3508	1.85	1.15
-2	480	.0495	.4036	.3992	1.51	0.95
-1	480	.0508	.4515	.4492	1.23	0.77
0	481	.0512	.5000	.5000	1	0.63
1	480	.0508	.5485	.5508	1.23	0.77
2	480	.0495	.5964	.6008	1.51	0.95
3	478	.0476	.6429	.6492	1.85	1.15
4	477	.0449	.6875	.6954	2.28	1.4
5	474	.0418	.7297	.7387	2.83	1.69
6	472	.0382	.7690	.7787	3.52	2.05
7	468	.0343	.8050	.8150	4.4	2.48
8	465	.0304	.8376	.8474	5.55	2.99
9	460	.0264	.8665	.8758	7.05	3.57
10	456	.0226	.8919	.9004	9.04	4.27
11	450	.0190	.9136	.9213	11.7	5.15
12	445	.0157	.9321	.9387	15.31	6.13
13	438	.0128	.9474	.9529	20.25	7.15
14	432	.0102	.9598	.9645	27.14	8.48
15	424	.0080	.9699	.9736	36.84	9.86
16	417	.0062	.9778	.9807	50.72	11.12
17	408	.0047	.9838	.9861	70.88	12.03
18	400	.0035	.9885	.9902	100.6	13.93
19	390	.0025	.9919	.9932	145.1	13.75
20	381	.0018	.9944	.9953	212.8	15.18
21	370	.0013	.9962	.9969	317.7	16.38
22	360	.0009*	.9975	.9979	483	17.34
* extreme values omitted						

Appendix G.

Multinomial Reference Distribution of ESS Grand Totals with 5 Repetitions of 4 Relevant Questions

score	ways	pmf	cdf	cdfContCor	odds	oddsLL05
-25	684	.0009*	.0029	.0024	407.7	21.86
-24	697	.0012	.0041	.0035	286.3	20.43
-23	708	.0016	.0057	.0049	203.6	18.75
-22	720	.0022	.0078	.0068	146.6	19.36
-21	730	.0028	.0107	.0093	106.9	17.1
-20	741	.0037	.0143	.0125	78.83	16.24
-19	750	.0047	.0190	.0167	58.81	13.85
-18	760	.0059	.0248	.0221	44.36	12.46
-17	768	.0074	.0321	.0287	33.81	10.92
-16	777	.0091	.0411	.0370	26.04	9.4
-15	784	.0110	.0519	.0471	20.24	8.24
-14	792	.0132	.0649	.0592	15.88	7.08
-13	798	.0156	.0801	.0737	12.57	6.01
-12	805	.0183	.0978	.0907	10.02	5.14
-11	810	.0211	.1182	.1104	8.06	4.36
-10	816	.0240	.1413	.1330	6.52	3.71
-9	820	.0270	.1671	.1585	5.31	3.14
-8	825	.0300	.1957	.1870	4.35	2.65
-7	828	.0329	.2270	.2185	3.58	2.24
-6	832	.0356	.2608	.2527	2.96	1.9
-5	834	.0381	.2968	.2895	2.45	1.6
-4	837	.0402	.3349	.3286	2.04	1.35
-3	838	.0420	.3746	.3697	1.71	1.14
-2	840	.0433	.4157	.4123	1.43	0.96
-1	840	.0441	.4576	.4559	1.19	0.81
0	841	.0444	.5000	.5000	1	0.68
1	840	.0441	.5424	.5441	1.19	0.81
2	840	.0433	.5843	.5877	1.43	0.96
3	838	.0420	.6254	.6303	1.71	1.14
4	837	.0402	.6651	.6714	2.04	1.35
5	834	.0381	.7032	.7105	2.45	1.6
6	832	.0356	.7392	.7473	2.96	1.9
7	828	.0329	.7730	.7815	3.58	2.24
8	825	.0300	.8043	.8130	4.35	2.65
9	820	.0270	.8329	.8415	5.31	3.14
10	816	.0240	.8587	.8670	6.52	3.71
11	810	.0211	.8818	.8896	8.06	4.36
12	805	.0183	.9022	.9093	10.02	5.14
13	798	.0156	.9199	.9263	12.57	6.01
14	792	.0132	.9351	.9408	15.88	7.08
15	784	.0110	.9481	.9529	20.24	8.24
16	777	.0091	.9589	.9630	26.04	9.4
17	768	.0074	.9679	.9713	33.81	10.92
18	760	.0059	.9752	.9780	44.36	12.46
19	750	.0047	.9810	.9833	58.81	13.85
20	741	.0037	.9857	.9875	78.83	16.24
21	730	.0028	.9894	.9907	106.9	17.1
22	720	.0022	.9922	.9932	146.6	19.36
23	708	.0016	.9943	.9951	203.6	18.75
24	697	.0012	.9959	.9965	286.3	20.43
25	684	.0009*	.9971	.9976	407.7	21.86

* extreme values omitted

Appendix H.

Multinomial Reference Distribution of ESS Subtotals with 5 Repetitions

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-14	16	.0005*	.0007	.0005	1970	44.38	12.54	6.66	6.11	4.19	2.85	2.1
-13	20	.0011	.0018	.0013	778.5	27.9	9.2	5.28	6.01	4	2.46	1.8
-12	25	.0022	.0040	.0029	339.5	18.43	6.98	4.29	5.82	3.56	2.17	1.55
-11	30	.0042	.0082	.0062	161.1	12.69	5.44	3.56	5.46	2.87	1.84	1.35
-10	36	.0074	.0156	.0120	82.2	9.07	4.35	3.01	4.92	2.44	1.57	1.18
-9	40	.0122	.0275	.0219	44.7	6.69	3.55	2.59	4.2	2.11	1.34	1.05
-8	45	.0188	.0458	.0375	25.68	5.07	2.95	2.25	3.86	1.74	1.17	0.93
-7	48	.0272	.0719	.0607	15.48	3.94	2.49	1.98	3.23	1.47	1.02	0.83
-6	52	.0374	.1072	.0933	9.72	3.12	2.13	1.77	2.56	1.22	0.89	0.75
-5	54	.0487	.1524	.1367	6.32	2.51	1.85	1.59	2.02	1.02	0.78	0.68
-4	57	.0602	.2075	.1914	4.23	2.06	1.62	1.43	1.53	0.86	0.69	0.62
-3	58	.0710	.2717	.2571	2.89	1.7	1.42	1.3	1.15	0.72	0.61	0.56
-2	60	.0798	.3434	.3322	2.01	1.42	1.26	1.19	0.84	0.61	0.54	0.51
-1	60	.0855	.4203	.4143	1.41	1.19	1.12	1.09	0.61	0.51	0.48	0.47
0	61	.0875	.5000	.5000	1	1	1	1	0.43	0.43	0.43	0.43
1	60	.0855	.5797	.5857	1.41	2	2.83	4	0.61	0.84	1.13	1.49
2	60	.0798	.6566	.6678	2.01	4.04	8.12	16.32	0.84	1.47	2.35	3.33
3	58	.0710	.7283	.7429	2.89	8.35	24.13	69.71	1.15	2.4	3.75	4.75
4	57	.0602	.7925	.8086	4.23	17.85	75.4	318.5	1.53	3.5	4.83	5.79
5	54	.0487	.8476	.8633	6.32	39.91	252.2	1593	2.02	4.05	5.7	6.1
6	52	.0374	.8928	.9067	9.72	94.48	918.4	8927	2.56	5.05	6.04	6.16
7	48	.0272	.9281	.9393	15.48	239.6	3710	57430	3.23	5.68	6.14	6.17
8	45	.0188	.9542	.9625	25.68	659.7	16940	435200	3.86	5.99	6.17	6.18
9	40	.0122	.9725	.9781	44.7	1998	89300	3991000	4.2	6.11	6.18	6.18
10	36	.0074	.9844	.9880	82.2	6756	555300	4.57E+07	4.92	6.16	6.18	6.18
11	30	.0042	.9918	.9938	161.1	25940	4178000	6.73E+08	5.46	6.17	6.18	6.18
12	25	.0022	.9960	.9971	339.5	115300	3.91E+07	1.33E+10	5.82	6.18	6.18	6.18
13	20	.0011	.9982	.9987	778.5	606100	4.72E+08	3.67E+11	6.01	6.18	6.18	6.18
14	16	.0005*	.9993	.9995	1970	3.88E+06	7.64E+09	1.51E+13	6.11	6.18	6.18	6.18

* extreme values omitted

Appendix I.

Multinomial Reference Distribution for ESS Grand Totals with 5 Repetitions of 2 Relevant Questions with PLE Sensor

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-20	2481	.0008*	.0019	.0015	659.2	15.03
-19	2645	.0012	.0031	.0025	402.9	14.37
-18	2808	.0018	.0048	.0039	252.4	13.47
-17	2967	.0026	.0074	.0061	161.9	12.31
-16	3123	.0038	.0112	.0093	106.2	10.94
-15	3273	.0053	.0164	.0139	71.17	10.8
-14	3418	.0072	.0235	.0201	48.68	9.9
-13	3555	.0097	.0331	.0286	33.94	8.1
-12	3685	.0127	.0455	.0398	24.1	6.95
-11	3805	.0162	.0613	.0543	17.4	6.03
-10	3916	.0203	.0809	.0727	12.76	5.05
-9	4015	.0248	.1047	.0953	9.49	4.14
-8	4105	.0297	.1330	.1226	7.15	3.4
-7	4183	.0347	.1659	.1549	5.45	2.77
-6	4252	.0398	.2034	.1923	4.2	2.25
-5	4309	.0447	.2452	.2346	3.26	1.82
-4	4357	.0491	.2910	.2815	2.55	1.47
-3	4393	.0528	.3401	.3323	2.01	1.18
-2	4420	.0556	.3919	.3863	1.59	0.95
-1	4435	.0574	.4455	.4426	1.26	0.76
0	4441	.0580	.5000	.5000	1	0.6
1	4435	.0574	.5545	.5574	1.26	0.76
2	4420	.0556	.6081	.6137	1.59	0.95
3	4393	.0528	.6599	.6677	2.01	1.18
4	4357	.0491	.7090	.7185	2.55	1.47
5	4309	.0447	.7548	.7654	3.26	1.82
6	4252	.0398	.7966	.8077	4.2	2.25
7	4183	.0347	.8341	.8451	5.45	2.77
8	4105	.0297	.8670	.8774	7.15	3.4
9	4015	.0248	.8953	.9047	9.49	4.14
10	3916	.0203	.9191	.9273	12.76	5.05
11	3805	.0162	.9387	.9457	17.4	6.03
12	3685	.0127	.9545	.9602	24.1	6.95
13	3555	.0097	.9669	.9714	33.94	8.1
14	3418	.0072	.9765	.9799	48.68	9.9
15	3273	.0053	.9836	.9861	71.17	10.8
16	3123	.0038	.9888	.9907	106.2	10.94
17	2967	.0026	.9926	.9939	161.9	12.31
18	2808	.0018	.9952	.9961	252.4	13.47
19	2645	.0012	.9969	.9975	402.9	14.37
20	2481	.0008*	.9981	.9985	659.2	15.03
* extreme values omitted						

Appendix J.

Multinomial Reference Distribution for ESS Grand Totals with 5 Repetitions of 3 Relevant Questions with PLE Sensor

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-24	9915	.0008*	.0023	.0019	518.7	21.4
-23	10248	.0011	.0034	.0028	352.2	20.18
-22	10572	.0015	.0048	.0041	242.7	18.69
-21	10888	.0020	.0069	.0059	169.7	16.95
-20	11193	.0027	.0096	.0082	120.4	17.25
-19	11488	.0036	.0132	.0114	86.55	14.98
-18	11770	.0047	.0179	.0156	63.05	13.98
-17	12040	.0061	.0239	.0210	46.52	12.51
-16	12295	.0077	.0316	.0280	34.75	10.89
-15	12536	.0097	.0411	.0367	26.26	9.29
-14	12760	.0119	.0527	.0475	20.06	8.05
-13	12970	.0144	.0668	.0607	15.49	6.83
-12	13163	.0172	.0835	.0765	12.07	5.74
-11	13342	.0202	.1031	.0953	9.5	4.85
-10	13504	.0235	.1257	.1172	7.53	4.06
-9	13652	.0269	.1514	.1424	6.02	3.41
-8	13783	.0303	.1803	.1710	4.85	2.85
-7	13900	.0336	.2122	.2030	3.93	2.39
-6	14000	.0369	.2471	.2383	3.2	2
-5	14086	.0398	.2847	.2766	2.62	1.67
-4	14155	.0424	.3247	.3177	2.15	1.39
-3	14210	.0446	.3667	.3611	1.77	1.16
-2	14248	.0461	.4102	.4064	1.46	0.97
-1	14272	.0471	.4548	.4529	1.21	0.8
0	14279	.0475	.5000	.5000	1	0.67
1	14272	.0471	.5452	.5471	1.21	0.8
2	14248	.0461	.5898	.5936	1.46	0.97
3	14210	.0446	.6333	.6389	1.77	1.16
4	14155	.0424	.6753	.6823	2.15	1.39
5	14086	.0398	.7153	.7234	2.62	1.67
6	14000	.0369	.7529	.7617	3.2	2
7	13900	.0336	.7878	.7970	3.93	2.39
8	13783	.0303	.8197	.8290	4.85	2.85
9	13652	.0269	.8486	.8576	6.02	3.41
10	13504	.0235	.8743	.8828	7.53	4.06
11	13342	.0202	.8969	.9047	9.5	4.85
12	13163	.0172	.9165	.9235	12.07	5.74
13	12970	.0144	.9332	.9393	15.49	6.83
14	12760	.0119	.9473	.9525	20.06	8.05
15	12536	.0097	.9590	.9633	26.26	9.29
16	12295	.0077	.9685	.9720	34.75	10.89
17	12040	.0061	.9761	.9790	46.52	12.51
18	11770	.0047	.9821	.9844	63.05	13.98
19	11488	.0036	.9868	.9886	86.55	14.98
20	11193	.0027	.9904	.9918	120.4	17.25
21	10888	.0020	.9931	.9941	169.7	16.95
22	10572	.0015	.9952	.9959	242.7	18.69
23	10248	.0011	.9966	.9972	352.2	20.18
24	9915	.0008*	.9977	.9981	518.7	21.4
* extreme values omitted						

Appendix K.

Multinomial Reference Distribution for ESS Grand Totals with 5 Repetitions of 4 Relevant Questions with PLE Sensor

score	ways	*pmf*	*cdf*	cdfContCor	odds	oddsLL05
-27	25602	.0008*	.0029	.0025	401.4	26.27
-26	26128	.0011	.0040	.0034	290	24.39
-25	26638	.0014	.0054	.0047	211.8	22.26
-24	27133	.0019	.0073	.0064	156.3	22.92
-23	27610	.0024	.0097	.0085	116.5	20.21
-22	28070	.0031	.0128	.0113	87.72	19.23
-21	28510	.0039	.0167	.0148	66.68	16.45
-20	28931	.0049	.0215	.0192	51.17	14.88
-19	29330	.0060	.0274	.0246	39.62	13.13
-18	29710	.0073	.0347	.0313	30.95	11.81
-17	30068	.0088	.0434	.0394	24.38	10.08
-16	30407	.0106	.0538	.0491	19.36	8.74
-15	30724	.0125	.0660	.0607	15.49	7.62
-14	31022	.0145	.0802	.0742	12.48	6.47
-13	31298	.0168	.0965	.0899	10.12	5.54
-12	31555	.0192	.1151	.1079	8.27	4.76
-11	31790	.0217	.1360	.1284	6.79	4.06
-10	32006	.0242	.1594	.1514	5.61	3.48
-9	32200	.0268	.1851	.1770	4.65	2.97
-8	32375	.0293	.2131	.2051	3.88	2.54
-7	32528	.0318	.2434	.2356	3.24	2.16
-6	32662	.0340	.2757	.2685	2.72	1.85
-5	32774	.0361	.3100	.3035	2.3	1.58
-4	32867	.0378	.3459	.3404	1.94	1.35
-3	32938	.0392	.3832	.3789	1.64	1.15
-2	32990	.0403	.4215	.4185	1.39	0.98
-1	33020	.0409	.4606	.4591	1.18	0.83
0	33031	.0411	.5000	.5000	1	0.71
1	33020	.0409	.5394	.5409	1.18	0.83
2	32990	.0403	.5785	.5815	1.39	0.98
3	32938	.0392	.6168	.6211	1.64	1.15
4	32867	.0378	.6541	.6596	1.94	1.35
5	32774	.0361	.6900	.6965	2.3	1.58
6	32662	.0340	.7243	.7315	2.72	1.85
7	32528	.0318	.7566	.7644	3.24	2.16
8	32375	.0293	.7869	.7949	3.88	2.54
9	32200	.0268	.8149	.8230	4.65	2.97
10	32006	.0242	.8406	.8486	5.61	3.48
11	31790	.0217	.8640	.8716	6.79	4.06
12	31555	.0192	.8849	.8921	8.27	4.76
13	31298	.0168	.9035	.9101	10.12	5.54
14	31022	.0145	.9198	.9258	12.48	6.47
15	30724	.0125	.9340	.9393	15.49	7.62
16	30407	.0106	.9462	.9509	19.36	8.74
17	30068	.0088	.9566	.9606	24.38	10.08
18	29710	.0073	.9653	.9687	30.95	11.81
19	29330	.0060	.9726	.9754	39.62	13.13
20	28931	.0049	.9785	.9808	51.17	14.88
21	28510	.0039	.9834	.9852	66.68	16.45
22	28070	.0031	.9872	.9887	87.72	19.23
23	27610	.0024	.9903	.9915	116.5	20.21
24	27133	.0019	.9927	.9936	156.3	22.92
25	26638	.0014	.9946	.9953	211.8	22.26
26	26128	.0011	.9960	.9966	290	24.39
27	25602	.0008*	.9971	.9975	401.4	26.27

* extreme values omitted

Appendix L.

Multinomial Reference Distribution of ESS Subtotals with 5 Repetitions with PLE Sensor

score	ways	*pmf*	*cdf*	Cdf ContCor	odds	odds 2RQs	odds 3RQs	odds 4RQs	odds LL05	odds2RQLL05	odds3RQLL05	odds4RQLL05
-15	161	.0005*	.0009	.0007	1517	38.94	11.49	6.24	7.71	5.36	3.32	2.24
-14	200	.0011	.0020	.0015	682.2	26.12	8.8	5.11	7.56	4.48	2.84	1.98
-13	243	.0021	.0041	.0030	328.4	18.12	6.9	4.26	7.27	4	2.42	1.73
-12	287	.0037	.0077	.0059	168	12.96	5.52	3.6	6.79	3.44	2.07	1.51
-11	333	.0062	.0139	.0109	90.88	9.53	4.5	3.09	6.1	2.91	1.81	1.35
-10	378	.0099	.0236	.0190	51.67	7.19	3.73	2.68	5.22	2.5	1.56	1.2
-9	423	.0150	.0383	.0315	30.72	5.54	3.13	2.35	4.84	2.08	1.37	1.07
-8	465	.0216	.0592	.0500	19.01	4.36	2.67	2.09	4.11	1.76	1.19	0.96
-7	505	.0297	.0875	.0758	12.19	3.49	2.3	1.87	3.3	1.49	1.05	0.87
-6	540	.0389	.1242	.1104	8.06	2.84	2.01	1.69	2.66	1.26	0.93	0.79
-5	571	.0489	.1697	.1546	5.47	2.34	1.76	1.53	2.06	1.06	0.83	0.72
-4	595	.0588	.2236	.2087	3.79	1.95	1.56	1.4	1.58	0.9	0.74	0.66
-3	615	.0678	.2852	.2720	2.68	1.64	1.39	1.28	1.19	0.77	0.66	0.61
-2	628	.0750	.3531	.3432	1.91	1.38	1.24	1.18	0.89	0.66	0.59	0.56
-1	637	.0797	.4254	.4201	1.38	1.18	1.11	1.08	0.65	0.56	0.53	0.52
0	639	.0814	.5000	.5000	1	1	1	1	0.48	0.48	0.48	0.48
1	637	.0797	.5746	.5799	1.38	1.91	2.63	3.63	0.65	0.89	1.18	1.52
2	628	.0750	.6469	.6568	1.91	3.66	7.01	13.41	0.89	1.54	2.45	3.52
3	615	.0678	.7148	.7280	2.68	7.16	19.17	51.3	1.19	2.5	4.13	5.21
4	595	.0588	.7764	.7913	3.79	14.38	54.52	206.7	1.58	3.68	5.31	6.97
5	571	.0489	.8303	.8454	5.47	29.89	163.4	893.3	2.06	4.78	6.77	7.62
6	540	.0389	.8758	.8896	8.06	64.9	522.8	4212	2.66	5.6	7.47	7.8
7	505	.0297	.9125	.9242	12.19	148.5	1810	22060	3.3	6.68	7.73	7.84
8	465	.0216	.9408	.9500	19.01	361.4	6870	130600	4.11	7.32	7.82	7.84
9	423	.0150	.9617	.9685	30.72	943.7	28990	890600	4.84	7.64	7.84	7.85
10	378	.0099	.9764	.9810	51.67	2669	137900	7126000	5.22	7.77	7.84	7.85
11	333	.0062	.9861	.9891	90.88	8259	750600	6.82E+07	6.1	7.82	7.85	7.85
12	287	.0037	.9923	.9941	168	28240	4745000	7.97E+08	6.79	7.84	7.85	7.85
13	243	.0021	.9959	.9970	328.4	107800	3.54E+07	1.16E+10	7.27	7.84	7.85	7.85
14	200	.0011	.9980	.9985	682.2	465300	3.17E+08	2.17E+11	7.56	7.85	7.85	7.85
15	161	.0005*	.9991	.9993	1517	2.30E+06	3.49E+09	5.29E+12	7.71	7.85	7.85	7.85

* extreme values omitted