Using the Empirical Scoring System

Raymond Nelson, Mark Handler, Pam Shaw, Michael Gougler, Benjamin Blalock, Chad Russell, Barry Cushman, and Marty Oelrich*

Abstract
The authors describe the procedures of the Empirical Scoring System, an evidence-based normative system for test data analysis of psychophysiological detection of deception examination data from comparison question test formats. A procedural description is provided for all aspects of the scoring model, including physiological features, mathematical transformations, decision rules, and cutscores based on normative data. Validation data for 13 dimensions of criterion accuracy are shown for 5,192 scored results of 732 confirmed Federal and Utah three-question ZCT examinations. These data are from six samples, scored by 140 experienced and inexperienced scorers in 16 cohorts of study participants. Normative data is provided in the appendices.

(The scene is a courtroom and you are the examiner answering questions as an expert witness.)

OPPOSING COUNSEL: What is the level of statistical significance or the probability of error for your manually scored polygraph examination result?

EXAMINER: Excuse me, can you repeat the question? I don't quite understand...

OPPOSING COUNSEL: What is the p-value or probability of error associated with your numerically scored test result?

EXAMINER: P-value? I don't know...

(To be continued...)

Introduction

The Empirical Scoring System (ESS) is an evidence-based numerical scoring model for manual test data analysis (TDA) of psychophysiological detection of deception (PDD) test data from examinations conducted using comparison question test (CQT) formats. The ESS is designed to provide the field examiner with a validated and reliable scoring model based on the simplest available solutions for which there is published evidence of validity to support all principles, assumptions and procedures. The simplicity of the ESS means that scoring skills are readily acquired by inexperienced scorers, easily retained by both experienced and inexperienced scorers, and easily understood when explained to non-polygraph professionals.

The ESS includes a description of the physiological data features that are correlated with truth and deception, mathematical transformation methods for assigning and aggregating numerical scores, decision rules

* Please contact the first author at raymond.nelson@gmail.com with any questions or comments. The authors grant unlimited use and duplication rights to any polygraph school accredited by the American Polygraph Association or recognized by the American Association of Police Polygraphists or the National Polygraph Association.

for the classification of numerical scores as indicative of truthfulness or deception, and numerical cutscores that define the a priori thresholds of statistical significance[1]. ESS cutscores are based on normative data that allow for calculation of the probability of an erroneous test result. Thus, the ESS allows for the selection of statistically optimal cutscores based on operational needs for the resolution and precision of the test result.

The ESS can be used with a variety of CQT formats that conform to valid principles for the construction of event-specific single-issue and multi-facet examinations, or multiple-issue screening exams. Applicable polygraph techniques include diagnostic test formats using three-question and two-question event-specific formats from the family of Zone Comparison Techniques (ZCT), multiple-facet investigative test formats based on the family of Modified General Question Techniques (MGQT), and multiple-issue screening formats whether based on the family of MGQTs or the Directed Lie Screening Techniques (DLST) (originally developed as the Test for Espionage and Sabotage [TES]). The ESS can be used with probable lie or directed lie comparison questions.

ESS Procedures

TDA begins with a global analysis of the data in order to ensure a sufficient quantity of data that is of satisfactory interpretable quality. Global analysis involves the inspection of the data for artifacts or events which may affect the interpretable quality of the physiological data in any of the recorded channels. Artifacts and uninterpretable events may include inattention and non-response to the test stimuli, indicators of behavioral non-cooperation, timeliness of reactions, voluntary or involuntary movements, and physiological responses to stimuli that are external to the examination. Global analysis should also attend to the pre-stimulus segment for all test questions. Reactions that occur in response to questions presented during artifacted or unstable segments should not be scored.

Features

The ESS requires no mechanical measurements and uses only those physiological signals that have been historically considered primary. The ESS does not include physiological signals that are considered secondary. When using the ESS, data are scored on screen through visual analysis, without the use of printing or mechanical measurements. A fixed measurement period is not used with the ESS. Instead, responses are scored until the end of the response whenever the onset of response is arguably attributable to the test stimuli, as determined by the overall quality and stability of the data and timeliness of the response onset. Responses are not scored when there is an apparent alternative cause, other than the test stimulus. Responses that begin before the onset of the stimulus and responses that begin more than several seconds after the answer cannot be reliably attributed to the test stimulus and are therefore not scored.

Electrodermal activity (EDA) is observed as the vertical rise of amplitude from the lowest point following the stimulus onset until the end of the reaction. Cardiovascular (CV) activity is observed as the vertical increase of amplitude at the diastolic baseline from the lowest point following the stimulus onset until the end of the reaction.

ESS is based on visual pattern recognition instead of physical or linear measurements and employs a simplified set of respiratory reaction patterns, involving reduction of respiration amplitude, slowing of respiration rate, and temporary increase of respiratory baseline for three cycles or more. Apnea, which is strongly correlated with the criterion of truth or deception, and has been described as the ultimate form of respiratory suppression, is easily feigned and should be scored only when it occurs at the relevant questions.

[1] Significance, for ESS scores, refers to the concept of statistical significance, or probability that a numerical test result has occurred due to error or random chance alone.

Transformations

The ESS employs a three position scale (+, 0, -), and scores are assigned using the *bigger-is-better* rule. Scores are assigned whenever there exists a visibly perceptible difference in the strength of reaction between relevant and comparison questions. This process is based on the simple and robust principle that physiological reactions of greater magnitude are caused by stimuli that are more salient to the examinee due to emotional, cognitive, or behaviorally conditioned factors. Positive (+) scores are assigned when there is a larger response to the comparison stimuli. Negative (-) scores when there is a larger response to the relevant stimuli. Relevant questions should be compared to comparison questions normally called for by the technique used.

The ESS is designed to place 50% of the diagnostic weight on the EDA by doubling all EDA scores. All EDA reactions are assigned a score of +2, 0 or -2, and all pneumograph and cardiograph reactions are assigned the values of +1, 0 and -1, regardless of the size difference between reactions. A single score is assigned for the combined upper and lower pneumograph sensors. This weighting scheme results in a reduction of inconclusive results, an increase in test sensitivity to deception, and no change in test specificity to truthfulness.

Decision Rules

Decision rules and cutscores are used together to interpret the numerical scores into categorical test results. Cut scores are based on normative data, and are selected for their desired level of statistical significance and tolerance for error with deceptive and truthful results. Decision rules are the procedural steps for interpreting the transformed and aggregated numerical scores into categorical results (i.e., truthful or deceptive) and are based on established tolerances for different types of errors. Different decision rules and cutscores can be used with the ESS to achieve different operational objectives such as test sensitivity, balanced test sensitivity and test specificity, or overall decision accuracy.

Grand Total Rule. The Grand Total Rule is the simplest of all decision rules and often provides the highest level of decision accuracy at the cost of a slightly higher inconclusive rate. The Grand Total Rule involves the summation of all scores into a single numerical result. Subtotal scores are not used with the Grand Total Rule.

1. If the grand total is less than or equal to the required cutscore for a statistically significant[2] deceptive result then the correct interpretation is Deception Indicated (DI).

2. If the grand total is greater than or equal to the required cutscore for a statistically significant truthful result, then the correct interpretation is No Deception Indicated (NDI).

3. If the grand total is not less than or equal to the required cutscore for a deceptive result, and it is not greater than or equal to the required cutscore for a truthful result, then the correct interpretation is Inconclusive (INC) or No Opinion (NO).

Spot-Score Rule. The Spot-Score Rule is used to interpret the results of multiple-issue screening exams and multiple-facet investigative exams conducted with formats from the family of MGQTs or DLST. This rule involves the calculation of subtotal scores for each relevant question. The Spot-Score Rule may be imposed when the Grand Total Rule does not lead to a definitive result in the case of single-issue testing, or as a stand-alone rule for multi-facet or multi-issue testing.

1. If *all* subtotal scores are greater than or equal to the required cutscore for a statistically significant truthful result, then the correct interpretation is No Significant Response (NSR) or NDI for

[2] Error rates for a range of cutscores have been calculated, and can be found in the appendices at the end of this article.

the test as a whole. Individual subtotal scores that are greater than or equal to the required cutscore for a statistically significant truthful result may be interpreted and reported as NSR (NDI) for the subtotal score and corresponding test question, if required by the referring agent or agency.

2. If *any* subtotal score is less than or equal to the required cutscore for a statistically significant deceptive result, then the correct interpretation is Significant Response (SR) or DI for the test as a whole as well as that subtotal.

3. When the test as a whole cannot be interpreted as SR (DI) or NSR (NDI), the overall test results should be interpreted as INC or NO. Do not interpret the results of any subtotal as NSR (NDI) if any subtotal has been interpreted as SR (DI). All remaining subtotals that are not SR (DI) are to be classified as INC or NO and the test as a whole is to be interpreted as SR (DI).

<u>Two-Stage Rules</u>. Two-Stage Rules function as the sequential use of the Grand Total Rule then Spot-Score Rules. Subtotal scores are not permitted to supersede the grand total score. Classifications based on grand total scores always trump the result of the subtotal scores. Two-Stage Rules have been shown to reduce the occurrence of inconclusive results and increase sensitivity to deception. Test results should be interpreted and reported at the level of the test-as-a-whole. However, the scored and interpreted results of individual questions may also be reported if required by the referring agent or agency.

<u>Stage One: Grand Total Rule</u>. Sum all scores for all questions to calculate a grand total score, representative of the test as a whole. Subtotal scores are not used during Stage One. Do not proceed to Stage Two if the grand total results in a DI or NDI.

1. If the grand total is greater than or equal to the required cutscore for a statistically significant truthful result, then the correct interpretation is NDI.

2. If the grand total is less than or equal to the required cutscore for a statistically significant deceptive result then the correct interpretation is DI.

3. Proceed to stage two only if the grand total results in INC or NO.

<u>Stage Two: Spot-Score Rule</u>. Grand total scores are not used in stage two. Do not interpret the results of some subtotals as SR (DI) and other subtotals as NSR (NDI) within a single examination.

1. If any subtotal score is less than or equal to the required cutscore for a statistically significant deceptive result, then the correct interpretation is DI for the test as a whole. Any subtotal score that is less than or equal to the required cutscore for a statistically significant deceptive result should be interpreted and reported as SR (DI) for the subtotal score and corresponding test question.

2. If none of the subtotal scores is less than or equal to the required cutscore for a statistically significant deceptive result, then the correct interpretation is INC or NO.

3. There is no solution to achieve a NDI result at stage two.

Cutscores and Normative Data

ESS cutscores are determined through the use of normative data, which are developed for different types of CQT examination formats. Cutscores are selected for their level of statistical significance according to requirements for test resolution or precision.

Three-question and two-question ZCT formats make use of the Grand Total Rule and may also employ the Spot-Score Rule at the second stage. Bonferonni correction is applied to the desired level of statistical significance when Spot-Score Rule is used with event-specific ZCT examinations[3].

More than 50% of truthful ZCT exams will include at least one subtotal score that is zero or lower. Ninety percent of all exams will produce a maximum difference of seven points or less, and 10% of exams will include a difference of eight or more points between subtotals. The effect of these concerns on decision accuracy is statistically insignificant.

Multiple-facet investigative exams, using MGQT formats, can also make use of both the Spot-Score Rule and the Grand Total Rule for improved criterion accuracy. Multiple-issue screening exams use only the Spot-Score Rule, and do not make use of the Grand Total Rule. Bonferonni correction is not applied to the cutscores for deceptive results of screening examinations. However, an inverse of the Šidák correction[4] is used to reduce inconclusive results among truthful cases while maintaining the required level of statistical significance.

Tables 1 and 2 show the cutscores for three-question and two-question exams from the ZCT family of examination formats. Table 3 shows the cutscores for multiple-issue and multiple-facet exams from the MGQT and DLST family of examination formats.

Appendices A and B show the normative data and lookup tables for ESS scores with three-question and two-question ZCT exams. Appendix C shows normative data and lookup data for multiple-facet and multiple-issue examinations with two, three, or four investigation targets. Normative data are based on a minimum of three presentations of each test stimulus question for which a minimum of two of three component scores from at least two of the three presentations must be interpretable, unless a particular technique includes more

Table 1. Cutscores for three-question ZCT exams.

	Cutscore	Alpha
NDI Grand Total	+2	< .10
NDI Grand Total (conservative)	+5	< .05
DI Grand Total	-4	< .05
DI Subtotal (when used)	-7 (any question)	< .017 (Bonferonni corrected)
NDI Subtotal (when used)	none	none

[3] Bonferonni is the name of a statistician who showed a simple procedure to correct for the potential for increased false-positive or type-1 errors that occur as a result of a phenomena know as *inflated alpha* which occurs in testing contexts for which statistical error probabilities are calculated repeatedly for a single issue of concern, such as when a polygraph test result will be decided at the level of the test as a whole while scoring the individual questions. For example if we set our alpha level to .05 then we have less than a 5% chance of error on each decision. If we make several decisions at this level of tolerance for error, the accumulated error rate will be compounded by the number of decisions (e.g., .05 * 3 relevant questions = .15). This could result in an error rate as high as 15% when we started with a declaration of tolerance of errors at no more than 5%. Bonferonni correction involved the selection of a cutscore that corresponds to the alpha level or tolerance for error divided by the number of decisions to be made (e.g., .05 / 3 relevant questions = Bonferonni corrected alpha = .0167). Bonferonni correction is often not used in screening contexts.

[4] Šidák is the name of a statistician who developed a mathematical correction to distortions of a desired alpha level, similar to the Bonferonni correction except that the Šidák correction assumes that the issues are independent and are being evaluated simultaneously within a single testing event. The inverse of this correction is used to correct the alpha and cutscores for truthful classification in multi-facet and multi-issue exams, when the questions are assumed to be independent and will be evaluated using the Spot-Score Rules which would result in a deflation of alpha and increased inconclusive results without this correction.

Table 2. Cutscores for two-question ZCT exams

	Cutscore	Alpha
NDI Grand Total	+2	< .10
NDI Grand Total (conservative)	+4	< .05
DI Grand Total	-4	< .05
DI Subtotal (when used)	-6	< .025 (Bonferonni corrected)
NDI Subtotal (when used)	none	none

Table 3. Cutscores for truthful results of multiple-issue and multiple-facet exams with two, three or four relevant questions

	Cutscore	Alpha
NDI Grand Total	none	none
DI Grand Total	none	none
DI Subtotal	-3 (any question)	< .05
NDI Subtotal	+1 (all questions)	< .10 (1-Šidák corrected)
NDI Subtotal (conservative)	+2 (all questions)	< .05 (1-Šidák corrected)

stringent requirements. Differences in normative distributions are statistically insignificant with additional presentations of the test question sequence, beyond the minimum of three presentations. It is therefore not necessary to change the normative cutscores when scoring a fourth of fifth presentation of the test stimuli.

Validation Data

Table 4 shows the criterion validity profile for ESS scores and four different decision rules for three-question ZCT examinations. Data are the scores (N = 5192) from 732 confirmed cases including Federal ZCT formats and, Utah ZCT collected from 16 cohorts of experienced and inexperienced scorers, totaling 140 participants, including field examiners, trainees, international trainees, and non-polygraph research psychologists who scored 6 different confirmed case samples.

The ESS achieves a high level of criterion accuracy, with mean criterion accuracy rates above 90% and mean inconclusives at approximately 10%.[5] The highest level of accuracy can be observed with the Grand Total Rule, with an inconclusive rate that is high though still under 20%. However, Two-Stage Rules will provide the optimal solution for most diagnostic and evidentiary purposes. Other decision rules will produce slightly different dimensional profiles for criterion accuracy. The Traditional ZCT Rule maintains a high level of decision accuracy, high negative predictive value, and low false-negative errors, though at a cost of higher inconclusives and weak specificity to truthfulness, compared to the other decision rules.

[5] Decision rules, together with the choice of cutscores, can affect the balance of test sensitivity and test specificity. Balanced sensitivity and specificity is desirable in evidentiary applications where false positives have a higher cost than they do in investigative applications. Investigative settings may require the avoidance of false-negatives, which may affect the balance of observed errors. What is optimum in one setting is not necessarily optimum in both settings.

Table 4. Criterion accuracy profile for three-question ZCT examinations

Decision Rules for ZCT Examinations Mean, (Standard Deviation) [95% CI]				
	Traditional ZCT Rules .10/.05 (+2tot/-4tot/ -7sub)	Two-stage Rules .10/.05 (+2tot/-4tot/-7sub)	Two-Stage Rules .05/.05 (+5tot/-4tot/ -7sub)	Grand Total Rule .05/.05 (+5tot/-4tot)
Correct	.930 (.032) [.867 to .993]	.922 (.028) [.867 to .977]	.935 (.027) [.882 to .987]	.947 (.025) [.899 to .996]
INC	.382 (.048) [.288 to .477]	.098 (.030) [.04 to .156]	.168 (.037) [.095 to .242]	.194 (.039) [.117 to .272]
D INC	.171 (.053) [.067 to .276]	.106 (.044) [.02 to .192]	.143 (.050) [.046 to .24]	.171 (.054) [.066 to .276]
T INC	.594 (.070) [.457 to .731]	.089 (.041) [.01 to .169]	.194 (.056) [.084 to .304]	.218 (.058) [.104 to .333]
Sensitivity	.822 (.054) [.715 to .928]	.817 (.055) [.709 to .924]	.820 (.054) [.713 to .927]	.792 (.058) [.679 to .905]
Specificity	.326 (.067) [.195 to .457]	.846 (.051) [.747 to .946]	.734 (.063) [.612 to .857]	.734 (.063) [.612 to .857]
FN	.007 (.012) [<.001 to .031]	.077 (.038) [.003 to .152]	.037 (.027) [<.001 to .089]	.037 (.027) [<.001 to .089]
FP	.08 (.038) [.005 to .154]	.064 (.034) [<.001 to .132]	.072 (.036) [.001 to .143]	.048 (.030) [<.001 to .107]
PPV	.912 (.042) [.830 to .993]	.927 (.039) [.852 to >.999]	.92 (.040) [.841 to .998]	.943 (.035) [.874 to 1.013]
NPV	.978 (.037) [.906 to >.999]	.916 (.041) [.835 to .997]	.952 (.035) [.884 to >.999]	.952 (.035) [.884 to >.999]
D Correct	.991 (.015) [.962 to >.999]	.913 (.042) [.831 to .996]	.957 (.031) [.896 to >.999]	.955 (.032) [.892 to >.999]
T Correct	.803 (.089) [.629 to .978]	.930 (.038) [.856 to >.999]	.911 (.045) [.823 to .999]	.939 (.038) [.864 to >.999]
Unweighted Average	.897 (.045) [.809 to .986]	.921 (.028) [.866 to .977]	.934 (.027) [.881 to .987]	.947 (.025) [.898 to .996]

Conclusions

The ESS is a simple and highly effective evidence-based TDA model. It is founded on decades of research into the principles that apply to both manual and automated test data analysis for CQT examinations. The advantages of a simple evidence-based scoring model are many, and can be expected to include: increased inter-scorer agreement compared to more complex models, accelerated skill acquisition among inexperienced scorers, increasing skill retention among experienced and inexperienced scorers, and increased generalizability of experimental results to field settings. Among the most important benefits of the ESS is it is based on normative data. PDD decision models based on normative data will allow field examiners to make inferential calculations of the probability of an erroneous test result. The ability to calculate and discuss a PDD test result in terms of a level of statistical significance or p-value (i.e., probability value or probability of error), as opposed to pass/fail categorical results, will mean that consumers and referring agents will be able to make better use of PDD examination results as a scientific decision-support tool that can add incremental validity to professional judgments regarding investigative priorities, interviewing objectives, risk assessment and risk management issues. The ESS presents an effective, expedient, and understandable evidence-based solution for field examiners, researchers, investigators, risk adjudicators, risk managers, judicial finders of fact and PDD program administrators.

The scene is a courtroom and you are the examiner answering questions as an expert witness.

(Take two)

OPPOSING COUNSEL: What is the level of statistical significance or probability of error for your manually scored polygraph examination result?

EXAMINER: Based on the normative data and the Empirical Scoring System, an evidence-based numerical scoring model for psychophysiological detection of deception test data analysis, the level of statistical significance or probability of error (p-value) for this test result is calculated at less than .001. In other words the probability that this test result was produced by a deceptive person (truthful person) is less than 1 in 1000.

OPPOSING COUNSEL: No further questions your honor!

References

Backster, C. (1963). Do the charts speak for themselves? New standards in polygraph chart interpretation. *Law and Order*, 11, 67-68, 71.

Barland, G. H. (1985). A method of estimating the accuracy of individual control question polygraph tests. In, Anti-terrorism; forensic science; psychology in police investigations: Proceedings of IDENTA-'85 (pp. 142-147). The International Congress on Techniques for Criminal Identification.

Blalock, B., Cushman, B. & Nelson, R. (2009). A replication and validation study on an empirically based manual scoring system. *Polygraph*, 38, 281-288.

Department of Defense Polygraph Institute (2006). Test Data Analysis: DoDPI numerical evaluation scoring system. Retrieved from http://www.antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf on 3-31-2007.

Handler, M., Nelson, R., Goodson, W. & Hicks, C. (2011). Empirical Scoring System: A Cross-cultural Replication and Extension Study of Manual Scoring and Decision Policies., *Polygraph*, 39, 200-215.

Handler, M., Nelson, R., Krapohl, D. & Honts, C. (2010). An EDA Primer for Polygraph Examiners. *Polygraph*, 32, 68-108.

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. SSD-PORPOR- 00-7272.

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

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.

Krapohl, D. J. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.

Nelson, R., Krapohl, D. J. & 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.

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.

Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37, 542-547.

Senter, S. M. (2003). Modified general question test decision rule exploration. *Polygraph*, 32, 251-263.

Senter, S. M. & Dollins, A. B. (2008). Exploration of a two-stage approach. *Polygraph*, 37, 149-164.

Van Herk, M. (1990). Numerical evaluation: Seven point scale +/-6 and possible alternatives: A discussion. *The Newsletter of the Canadian Association of Police Polygraphists*, 7, 28-47. Reprinted in *Polygraph*, 20, 70-79.

Appendix A

Normative Lookup Data for Three-Question ZCT Exams

Mean deceptive score = -9 (SD = 8)

Mean truthful score = 8 (SD = 7)

Three-question ZCT Truthful (NSR) Cut-scores	
Total NDI/NSR Cut-score	**p-value (alpha)**
-1	0.159
0	0.130
1	0.106
2	0.085
3	0.067
4	0.052
5	0.040
6	0.030
7	0.023
8	0.017
9	0.012
10	0.008
11	0.006
12	0.004
13	0.003
14	0.002
15	0.001
Three-question ZCT Deceptive (SR) Cut-scores	
Total DI/SR Cut-score	**p-value (alpha)**
1	0.159
0	0.127
-1	0.099
-2	0.077
-3	0.058
-4	0.043
-5	0.032
-6	0.023
-7	0.016
-8	0.011
-9	0.008
-10	0.005
-11	0.003
-12	0.002
-13	0.001

Appendix B

Normative Lookup Data for Two-Question ZCT Exams

Mean deceptive score = -6 (SD = 6)

Mean truthful score = 6 (SD = 6)

Two-question ZCT Truthful (NSR) Cut-scores	
Total NDI/NSR Cut-score	**p-value (alpha)**
-1	0.202
0	0.159
1	0.122
2	0.091
3	0.067
4	0.048
5	0.033
6	0.023
7	0.015
8	0.010
9	0.006
10	0.004
11	0.002
12	0.001
Two-question ZCT Deceptive (SR) Cut-scores	
Total DI/SR Cut-score	**p-value (alpha)**
1	0.202
0	0.159
-1	0.122
-2	0.091
-3	0.067
-4	0.048
-5	0.033
-6	0.023
-7	0.015
-8	0.010
-9	0.006
-10	0.004
-11	0.002
-12	0.001

Appendix C

Normative Lookup Data for Multiple-Issue and Multiple-Facet Exams

Mean deceptive score = -2 (SD = 3)

Mean truthful score = 2 (SD = 3)

Multiple-Issue and Multi-Facet Deceptive (SR) Cut-scores		
SR Cut-score	**p-value (alpha)**	
0	0.252	
-1	0.159	
-2	0.091	
-3	0.048	
-4	0.023	
-5	0.010	
-6	0.004	
-7	0.001	
Multiple-Issue and Multi-Facet Screening Truthful Cutscores		
NSR Cut-score (4RQs)	**Uncorrected Alpha**	**Observed Alpha**
0	0.252	0.070
1	0.159	0.042
2	0.091	0.024
3	0.048	0.012
4	0.023	0.006
5	0.010	0.002
6	0.004	0.001
Multiple-Issue and Multi-Facet Screening Truthful Cutscores		
NSR Cut-score (3RQs)	**Uncorrected Alpha**	**Observed Alpha**
0	0.252	0.092
1	0.159	0.056
2	0.091	0.031
3	0.048	0.016
4	0.023	0.008
5	0.010	0.003
6	0.004	0.001
Multiple-Issue and Multi-Facet Screening Truthful Cutscores		
NSR Cut-score (2RQs)	**Uncorrected Alpha**	**Observed Alpha**
0	0.252	0.135
1	0.159	0.083
2	0.091	0.047
3	0.048	0.024
4	0.023	0.011
5	0.010	0.005
6	0.004	0.002
7	0.001	0.001