



Marcin Gołaszewski*
Internal Security Agency
Warsaw
Poland

Validated Techniques and Scoring Models for PDD Test Data Analysis – Conclusions from the 2011 APA Report

Key Words: APA Report, standard of polygraph examination, polygraph techniques, recommended techniques

Standards concerning polygraph examinations for common application can be found in by-laws as well as in recommendations issued by the American Polygraph Association (APA) and in standards adopted by ASTM International (American Society for Testing and Materials International). The first of these organizations was founded in 1966 and has a membership of over three thousand polygraphists, whereas the second is a normalization organization which can trace its roots back to 1898. Both have the adjective “American” in their names, but in fact these are international organizations open to representatives from all over the world.¹

* marcin.gołaszewski@wp.pl

¹ The international character of the ASTM was underlined by augmenting the name with the element *international*. in 2001. A similar idea was an issue during the 45th annual APA seminar in

On 1 January 2012, the new **APA standards of practice** came into force, which introduced among other things:

- mandatory usage of a motion sensor for all examinations;
- obligatory polygraph instrument functionality test recorded semi-annually;
- general requirement for using only validated techniques (testing techniques shall be considered valid if supported by research conducted in accordance with the APA's research standards. For a minimum of five years after publication, upon request, researchers of polygraph techniques shall provide reasonable access to validation data for critical review. Where examinations deviate from the protocols of a validated testing technique, the deviations should be noted and justified in writing);
- criteria for the admissibility of particular techniques in specific types of examinations: evidentiary, paired-testing, investigative and screening.

As Pamela Shaw (APA President 2011-2012) rightly noted: “The requirement to use validated testing methods is not a new idea, of course. Other fields such as medicine and psychology eventually came to the same conclusion, albeit many years after the fields were established. It has turned out to be a great thing for them. Try to imagine, if you can, what the fields of medicine and psychology would be like if there were no requirement to validate their methods. Validation serves a number of important functions, not the least of which is protecting the public from misuse, incompetence and quackery” (APA, 2011).

It is worth specifying that the term *validus* in Latin means: strong, vigorous. By definition, method validation is the process of establishing the performance characteristics and limitations of a method. In polygraph testing **criterion validity refers to the ability of the test to correctly determine the truthful or deceptive criterion category to which an examination belongs**. It is important to determine whether the data analysis process according to a given method is **reliable** (a test will give the same result when the test is repeated or when the data are re-evaluated by another professional) and **generalizable** (a test that works on sample data will also work on other cases in the field). Validity is merely estimated from the published scientific reports.

Generalization of validity is not warranted when the structure or intended use of the test variant differs from a validated model to the extent that the

Myrtle Beach (2010). However, at that time the conservative approach prevailed – the argument concerning the recognizability of the previous brand that had been built up over many years.

distributions of scores can be expected to differ. For example: validation evidence for event-specific diagnostic techniques (interpreted with the assumption of non-independent criterion variance of the relevant questions) cannot be generalized to multi-issue screening variants of these techniques that are scored and interpreted with the assumption of independent criterion variance. Another example is when differences in the number of RQs affect the mean total score (APA, 2012).

It must be remembered that a **polygraph technique** is not just a test question sequence, but also a set of structured rules regarding: pre-test interview, target selection and question formulation, in-test stimuli presentation and test data analysis method. According to APA recommendations – in order to consider a given technique valid, it must be a combination of the following characteristics:

- test format that conforms to valid principles relating to: target selection, question formulation and in-test presentation of the stimulus questions,
- validated method for test data analysis (TDA),
- at least two studies (original and replication) published in: “Polygraph” or other peer reviewed journals, government publications or edited academic texts.

The APA also defined the **criteria that a technique authorized for different sorts of examinations should meet:**

- in **evidentiary testing** (commissioned by judicial bodies, prosecution, defence etc.): $\geq 90\%$ accuracy and $\leq 20\%$ inconclusive results,
- in **paired testing** (2 independent polygraphists examine at least 2 persons who testify in this way that one of them must surely be lying): $\geq 86\%$ accuracy and $\leq 20\%$ inconclusives rate,
- in **investigative testing**: $\geq 80\%$ accuracy and $\leq 20\%$ inconclusives,
- for **screening purposes**: an accuracy rate that is significantly greater than chance + successive hurdles approach which requires conducting additional validated and more precise tests if a screening test ends unfavourably (there are doubts regarding the examinee’s truthfulness).

Who is affected by these standards? Since 1 January 2012 – APA members. In case of standards violations (unless otherwise provided by state or national law), sanctions, including loss of membership, might be imposed. Another important organization – the AAPP (American Association of Police Polygraphists) – intends to adopt the same standards as of 2013. It is expected that other polygraph associations (in the U.S. and all over the world) may follow suit.

More rigorous standards that became effective recently had already been elaborated in 2007 as a response to the – essentially critical – 2002 **report on the polygraph by the National Research Council** of the USA. The findings and conclusions in this report were both positive and negative for the polygraph profession. However, the former predominated:

- the scientific basis for polygraph testing is far from desirable for a test that carries considerable weight in national security decision making;
- the bulk of polygraph research can accurately be characterized as atheoretical;
- basic psychophysiology gives cause for concern that effective countermeasures to the polygraph may be possible;
- available knowledge about the physiological responses measured by the polygraph suggests that there are serious upper limits in principle to the diagnostic accuracy of polygraph testing, even with advances in measurement and scoring techniques.

Fortunately, there were also findings justifying moderate optimism for the future. The NRC admitted that *although the basic science indicates that polygraph testing has inherent limits regarding its potential accuracy, it is possible for a test with such limits to attain sufficient accuracy to be useful in practical situations* (NRC, 2003). In the NRC meta-analysis, the range of accuracy rates for single issue tests was between 0.81 and 0.91 for the middle 26 values from 52 datasets. This means that – despite general severe criticism – the relatively high effectiveness of such examinations was confirmed. That was 10 years ago. Now we know enough to say a lot more.

The fundamental question is: **which of the PDD techniques that are well-known to date satisfy the new APA requirements?** An answer was given by a **meta-analysis** – an independent scientific study which relies on a secondary exploration of knowledge by means of a systematic review of the information contained in publications and original sources using: data connections, statistical analyses, generalization of results and inferences. The meta-analysis carried out by The APA Ad-Hoc Committee on Validated Techniques encompassed: 37 studies (52 experiments and surveys), 289 scorers, 12665 scored results of 4283 confirmed exams (6597 scored results of 2300 confirmed deceptive exams and 6068 scored results of 1983 confirmed truthful exams). Studies were weighted by sample size and number of participant scorers. As a result researchers evaluated:

- unweighted **accuracy of all recognizable PDD techniques** (without outliers) that produce generalizable results = 87.1%, with inconclusives rate: 12.7%,

- unweighted accuracy of **single-issue** techniques² (without outliers) = 92.1%, with INC: 8.8%,
- unweighted accuracy of **screening** techniques = 85%, with INC: 12.5%.

Moreover, the APA approved the **list of PDD validated techniques** suitable for use in specific types of examinations (see table 1). It has been effective since 1 January 2012.

Evidentiary techniques ³ / Test data analysis method	Paired testing techniques ⁴ / TDA method	Investigative techniques / TDA method
<p>Federal You-Phase / ESS⁵</p> <ul style="list-style-type: none"> • accuracy⁶: 90.4% • inconclusives (INC): 19.2% • sensitivity⁷: 84.5% • specificity⁸: 75.7% 	<p>AFMGQT⁹ / ESS</p> <ul style="list-style-type: none"> • accuracy: 87.5% • inconclusives (INC): 17% • sensitivity: 72.9% • specificity: 70% 	<p>AFMGQT / 7-position scale</p> <ul style="list-style-type: none"> • accuracy: 81.7% • inconclusives (INC): 19.7% • sensitivity: 78.3% • specificity: 53.8%

² In single-issue techniques the variance of response to individual questions is non-independent (affected by and/or affects the variance of response to other questions). In multi-faceted and multiple-issue techniques the criterion variance of the test questions is independent.

³ techniques used in exams to be admitted in court.

⁴ paired testing – a method of utilizing polygraph testing in situations in which two or more subjects give contradictory accounts of a particular incident in such a way that at least one of the subjects must certainly be lying. The method utilizes two independent examiners with established accuracy and error rates to assess the veracity of at least two subjects in such circumstances in which opposing parties assert diametrically opposed information as factual. See: *Model Policy for Paired Testing* [online], American Polygraph Association. Available from: http://www.polygraph.org/files/Model_Policy_for_Paired_Testing.doc [Accessed 13 November 2012].

⁵ Empirical Scoring System (ESS) – an evidence-based normative system for manual test data analysis of PDD examination data from comparison question test formats. For more details, see Nelson et al., 2011.

⁶ Accuracy – proportion of correct decisions, excluding inconclusives.

⁷ sensitivity – ability of a test to detect specific features at all levels of magnitude or prevalence. In PDD testing this term is used to describe how well a test identifies a person engaging in deception concerning the issue under investigation (Krapohl, Handler, Sturm, 2012). The proportion of true positives a test can produce.

⁸ specificity – the proportion of true negatives a test can produce. This term is used to describe how well a test identifies a person being truthful concerning the issue under investigation.

⁹ Two versions exist for the AFMGQT (1 and 2), with minor structural differences between them. Selected studies include a mixture of both AFMGQT versions, so these results are provided as generalizable to both versions. The two techniques are nearly identical to the **LEPET** and the **Utah MGQT**. That is why the validity of the AFMGQT can be generalized to these techniques if scored with the same TDA methods. Any hypothesis that the validity or criterion accuracy of AF MGQT and LEPET exams differs will require research evidence.

ZCT (Federal. Utah) / ESS <ul style="list-style-type: none"> • accuracy: 92.1% • INC: 9.8% • sensitivity: 81.7% • specificity: 84.6% 	Federal You-Phase / 7-pos. scale <ul style="list-style-type: none"> • accuracy: 88.3% • INC: 16.8% • sensitivity: 84.5% • specificity: 75.7% 	CIT (GKT) / Lykken system <ul style="list-style-type: none"> • accuracy: 82.3% • INC: 0.1% • sensitivity: 81.5% • specificity: 83.2%
Utah ZCT (combined versions) / Utah <ul style="list-style-type: none"> • accuracy: 93% • INC: 10.7% • sensitivity: 85.3% • specificity: 80.9% 	Federal ZCT / 7-pos. <ul style="list-style-type: none"> • accuracy: 86% • INC: 17.1% • sensitivity: 85.8% • specificity: 58.1% 	DLST (TES) / 7-pos. <ul style="list-style-type: none"> • accuracy: 84.4% • INC: 8.8% • sensitivity: 74.8% • specificity: 79.2%
Utah ZCT DLC / Utah <ul style="list-style-type: none"> • accuracy: 90.2% • INC: 7.3% • sensitivity: 81.5% • specificity: 85.7% 	Federal ZCT / 7-pos. evidentiary¹⁰ <ul style="list-style-type: none"> • accuracy: 88% • INC: 8.5% • sensitivity: 80.4% • specificity: 80.9% 	DLST (TES) / ESS <ul style="list-style-type: none"> • accuracy: 85.8% • INC: 9% • sensitivity: 80.9% • specificity: 75.1%
Utah ZCT PLC / Utah <ul style="list-style-type: none"> • accuracy: 93.1% • INC: 7.7% • sensitivity: 86.7% • specificity: 83.3% 	Backster You-Phase / Backster <ul style="list-style-type: none"> • accuracy: 86.2% • INC: 19.6% • sensitivity: 83.6% • specificity: 55.6% 	
Utah ZCT RCMP (v.1) / Utah <ul style="list-style-type: none"> • accuracy: 93.9% • INC: 18.5% • sensitivity: 83.3% • specificity: 70% 		
*IZCT / HSS <ul style="list-style-type: none"> • accuracy: 99.4% • INC: 3.3% • sensitivity: 97.7% • specificity: 94.6% 		

¹⁰ In the 7-position evidentiary scoring method the decision threshold for the opinion NDI is somewhat lower than in the traditional 7-pos. scale and amounts to +4. For the opinion DI it remains as previously (-6).

<p>*MQTZCT / Matte</p> <ul style="list-style-type: none"> • accuracy: 99.4% • INC: 2.9% • sensitivity: 96.7% • specificity: 96.3% 		
--	--	--

Table 1. The list of PDD validated techniques. Accuracy (correct decisions), inconclusive rates, sensitivity and specificity. (Content based on: *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, American Polygraph Association, 2011)

How to read the above table? Techniques that have $\geq 90\%$ accuracy and $\leq 20\%$ inconclusives were placed in the first column. In the second column – techniques with 86% accuracy and producing no more than 20% inconclusive results. And in the third - techniques with at least 80% accuracy and giving at most 20% inconclusives. Techniques from the first column can also be applied in examinations specified in the second column, while in investigative examinations one can use techniques mentioned in all the columns. Looking from the left to the right side of the table, criteria of admissibility (accuracy) become progressively lower.

In two techniques – the Integrated Zone Comparison Technique (IZCT) and the Matte Quadri-Track Zone Comparison Technique (MQTZCT) – comments were made on the references. These techniques have been listed in the table; however, it was indicated that statistical data are inconsistent with the distribution of results from all other techniques and are called **outliers**. Therefore one ought to look at these data with great caution. All the more so because the IZCT and the MQTZCT have not been verified by independent researchers. Furthermore, the APA drew attention to some shortcomings in the validation process of these techniques.

For example, the generalizability of results relating to IZCT is limited by the fact that no measures of test reliability have been published for this technique. There were also significant differences between sampling distributions from different studies.

Moreover, the developer of MQTZCT reported a near-perfect correlation coefficient of 0.99 for the numerical scores. He suggested an unprecedented high rate of inter-scorer agreement, which is unexpected bearing in mind the complexity of the method. In addition to this, scores were not provided for those cases that were not scored correctly.

Some popular techniques were omitted from the list. Among them one can mention: the US Army MGQT, Reid technique (GQT), searching POT, Marcy and R/I. The Army MGQT failed to satisfy criterion accuracy. Most studies regarding the Reid technique could not be included in the meta-analysis. The reasons for their exclusion include serious sampling confounds, insufficient information to calculate all of the statistics of interest to the meta-analysis, use of test-data-analysis models that differ substantially from the Reid method, and the use of instrumentation and testing procedures that differ substantially from actual field practices (APA, 2012). In turn, anyone using the R/I or Marcy techniques was permitted to do so throughout 2012 to allow time for further validation studies. However, there is no indication of Marcy's probable success. A few unpublished studies regarding the R/I technique exist, but they show only around a 75% accuracy level.

Examiners who want to use techniques researched by themselves should label such techniques as "experimental". Nothing precludes the use of supplementary techniques to support a decision based on a validated technique. However, such techniques shall not be used as the sole basis for a final opinion after a polygraph examination. The list of permissible polygraph techniques remains open. It can be extended provided that a technique fulfils criteria of scientific validation, minimum accuracy levels and maximum levels of inconclusive results.

The list of validated techniques includes **4 major standardized test data analysis methods**: 7-position US Federal, University of Utah, Empirical Scoring System and Lykken scoring.

Lykken scoring is the TDA method for the CIT/GKT. It entails the ranking of the electrodermal response amplitudes from 2 to 0. If the largest EDR takes place on the key item, the score for that test is a 2. If the second largest EDR takes place on the key item, the score is a 1. All others are scored 0. Reactions to the first buffer are ignored (Krapohl, McCloughan, Senter, 2006). The cutoff for a call of "RI – recognition indicated" is equal to the number of CIT subtests.

Other TDA methods are used for comparison question tests (see table 2 and 3). The oldest is the **US Federal Government scoring system** (a modification of the Backster scoring system developed in 1963). It was based on 22 diagnostic features taught by the United States Army Military Police School (Weaver, 1980). In 2006 the Department of Defense Polygraph Institute (DoDPI) made changes in physiological criteria (it kept 8 main features and introduced 3 auxiliary ones). Cutoff scores and decision rules were not modified at that

time. Since 2010, the National Center for Credibility Assessment (NCCA) has been responsible for the polygraph examinations program in the United States. Three variants of the Federal TDA model exist: “7-position”, “7-position evidentiary” and “3-position”. The first two are valid and satisfy APA 2012 standards. Decision accuracy for 3-position techniques was not significantly different from 7-position, but inconclusive rates were excessive and beyond the boundaries permitted by the APA 2012 standards. Nevertheless, the three-position scoring model is valid in a scientific sense and can be used in field settings when field practices require that the results of inconclusive tests are re-evaluated using another validated TDA model.

The next system was developed as a result of studies that had been carried out by researchers from the **University of Utah** (Salt Lake City) since the 1970s. The main researcher was David Raskin. They generally concluded that the numerical scoring of polygraph charts produces higher rates of accuracy and reliability than any sort of chart interpretation. However, they deemed systems known so far to be imperfect. Some elements of both existing models: the Backster and the US Army (in the version before the fundamental modification) did not have satisfactory scientific grounds. Researchers decided to modify the Backster system, which – in their opinion – contained too complicated rules and was disadvantageous for truthful persons. As a consequence, the complete Utah approach to comparison questions testing (including Utah ZCT, Utah MGQT) together with the numerical evaluation system were developed. These methods were confirmed by many research studies and peer reviewed publications in the following 30-40 years.

The newest, least complicated and also very well scientifically documented system is the so-called **Empirical Scoring System**. It was first described in 2008 by R. Nelson, M. Handler and D. Krapohl. Only main patterns of reactions from a wide group of diagnostic features described in the literature are subject to test data analysis in the ESS. Results of this analysis are compared to cutscores dependent on the adopted tolerance of error, the required level of statistical significance and the probability of error on the basis of representative data. Tolerance of error for deceptive scores was established at the 5% level ($\alpha = 0.05$), and for truthful results – at 10% ($\alpha = 0.1$). This concerns grand total scores. However, when decisions are made on the basis of subtotal scores, the Bonferroni correction is applied. This is a procedure to correct for the potential for increased false-positive errors. As a consequence, in ZCT formats with three relevant questions, alpha must be divided by 3 – that gives us corrected $\alpha = 0.017$.

Experiments have confirmed that ESS produces similar results when it is used both by qualified experts and inexperienced examiners. Therefore, it has a chance to become a main polygraph TDA model with universal application.

CHANNEL	TDA METHOD		
	US FEDERAL	UTAH	EMPIRICAL SCORING SYSTEM (ESS)
PNEUMO (respiration)	<p>Start of reaction: from the stimulus onset to 1 full cycle after the answer. Range of reaction: ≥ 3 cycles.</p> <ul style="list-style-type: none"> • suppression (decrease in amplitude), • apnea, • change in inhalation and exhalation ratio, • progressive decrease in amplitude, • slowing of rate, • temporary change in baseline (secondary feature – as contrasted with above – non-RLL feature). 	<p>Start of reaction: from the stimulus onset to 5 seconds after the answer. Range of reaction: ≥ 3 cycles, up to 20 seconds if response began at appropriate time.</p> <ul style="list-style-type: none"> • decrease in amplitude, • baseline arousal, • apnea, • slowing of rate. 	<p>Start of reaction: no rigid rules; generally from the stimulus onset to 5 sec. after the answer. Range of reaction: ≥ 3 cycles, up to 15-20 seconds.</p> <ul style="list-style-type: none"> • decrease in amplitude, • slowing of rate, • baseline arousal.
EDA (electrodermal activity)	<p>Start of reaction: from the stimulus onset to the answer. Range of reaction: from start of reaction to return to the baseline preceding stimulus onset.</p> <ul style="list-style-type: none"> • amplitude (main feature measured from the baseline to the peak of reaction), • complexity (the curve does not return to the baseline but another physiological arousal occurs), • duration (period of time between the start of reaction and return to the baseline). <p>The last two features are taken into account only when both compared EDA amplitudes are similar.</p>	<p>Start of reaction: 0.5 sec. from the stimulus onset to 5 sec. after the answer. Range of reaction: from start of reaction to return to the baseline.</p> <ul style="list-style-type: none"> • amplitude, • auxiliary: duration and complexity. 	<p>Start of reaction: no rigid rules; generally from the stimulus onset to 5 sec. after the answer. Range of reaction: up to 15-20 sec.</p> <ul style="list-style-type: none"> • amplitude.

<p>CARDIO (relative blood pressure and pulse rate)</p>	<p>Start of reaction: from the stimulus onset to the end of the answer. Range of reaction: from start of reaction to return to the baseline (on diastolic side).</p> <ul style="list-style-type: none"> • increase of baseline (main feature), • decrease in pulse rate (if the main feature does not occur), • duration (auxiliary – when compared changes of baseline are equal). 	<p>Start of reaction: from the stimulus onset to 5 seconds after the answer. Range of reaction: from start of reaction to the return to the baseline.</p> <ul style="list-style-type: none"> • baseline arousal (curve increase – more clear on diastolic side), • duration. 	<p>Start of reaction: from the stimulus onset to 5 seconds after the answer. Range of reaction: up to 15-20 seconds.</p> <ul style="list-style-type: none"> • amplitude (curve increase).
<p>PPG (changes in blood volume in blood vessels of the finger-tip of the hand)</p>	<ul style="list-style-type: none"> • no recommendations 	<p>Start of reaction: from the stimulus onset to 5 seconds after the answer. Range of reaction: up to 20 seconds.</p> <ul style="list-style-type: none"> • amplitude reduction and duration of that change. 	<p>Start of reaction: 2 seconds from the stimulus onset to 5 sec. after the answer. Range of reaction: up to 15-20 seconds.</p> <ul style="list-style-type: none"> • amplitude reduction.

Table 2. Diagnostic criteria used in validated polygraph test data analysis systems.

	TDA METHOD		
	US FEDERAL	UTAH	EMPIRICAL SCORING SYSTEM (ESS)
<p>GENERAL GUIDELINES</p>	<ul style="list-style-type: none"> • 7-position scale: 0 – equal or no responses to compare, +1/-1 – subtle difference, +2/-2 – definite difference, +3/-3 – dramatic difference. 	<ul style="list-style-type: none"> • 7-pos. scale: 0 – equal or no responses to compare +1/-1 – noticeable difference, +2/-2 – strong and clear difference, +3/-3 – dramatic difference, stable curve and the most significant response on the chart. • 3 or 5 charts (if inconclusive after 3 charts). 	<ul style="list-style-type: none"> • “bigger is better” rule – we score any noticeable difference between responses; • only 3-pos. scale [+1, 0, -1], ▪ exclusion for EDA: ▪ 3-pos. scale but scores are doubled: [+2, 0, -2].

PNEUMO	<ul style="list-style-type: none"> usually scores: +1/-1, very rarely +2/-2, never +3/-3. in case of two equivalent diagnostic features we measure the time window of longer reaction and then we compare length lines (RLL) in the same time windows of reactions. 	<ul style="list-style-type: none"> usually scores: +1/-1, very rarely +2/-2, never +3/-3. in case of two equivalent diagnostic features we take into account the duration of reactions (the segment of curve for comparison must be in the reaction window from stimulus onset to 10 subsequent seconds). 	<ul style="list-style-type: none"> reaction vs. reaction = 0 apnea is taken into consideration only at relevant questions (it's easy to create artificially), it is recommended to give 0 in the tests with directed lie questions.
EDA	<ul style="list-style-type: none"> 1 = amplitudes ratio not greater than 3:1, 2 = ratio > 3:1 < 4:1, 3 = ratio ≥ 4:1. if there is no reaction to one of the compared questions, we apply the rule regarding quantity of chart divisions: 1 = up to 2 divisions, 2 = from 2 to 3 divisions, 3 = more than 3 divisions. 	<ul style="list-style-type: none"> 1 = double difference in amplitude, or 1.5:1 ratio + duration and complexity, 2 = triple difference in amplitude, or 2.5:1 ratio + duration and complexity, 3 = quadruple difference in amplitude, and the most significant response on the chart. 	<ul style="list-style-type: none"> +2, 0, -2
CARDIO	<ul style="list-style-type: none"> 1 = up to 2 times greater increase in baseline, 2 = from 2 to 3 times greater reaction, 3 = at least 3 times greater reaction. if there is no reaction to one of the compared questions, we apply the rule regarding quantity of chart divisions: 1 = up to 2 divisions, 2 = from 2 to 3 divisions, 3 = more than 3 divisions. 	<ul style="list-style-type: none"> 1 = magnitudes of reactions ratio 1.5:1, 2 = ratio 2:1, 3 = ratio 3:1 and the most significant response on the chart. 	<ul style="list-style-type: none"> +1, 0, -1
PPG	<ul style="list-style-type: none"> no recommendations. 	<ul style="list-style-type: none"> scores 1 or 2, never 3. 	<ul style="list-style-type: none"> +1, 0, -1.

<p>CUT-OFF SCORES</p>	<ul style="list-style-type: none"> • the same cutscores for 7-pos. and 3-position scales. • ZCT DI – when grand total ≤ -6 or any subtotal ≤ -3 NDI – if every subtotal (spot total) ≥ +1 and grand total ≥ +6 INC – other results. • You-Phase (Bi-Zone) DI – when grand total ≤ -4 or any subtotal ≤ -3 NDI – if all subtotals ≥ +1 and grand total ≥ +4 INC – other results. • DLST SR – if grand total ≤ -4 or when any subtotal ≤ -3 NSR – when all spots ≥ +1 and grand total ≥ +4 INC – other results. 	<ul style="list-style-type: none"> • single issue test (Utah ZCT): DI – if grand total ≤ -6 NDI – grand total ≥ +6 INC – other results. • multi-faceted (Utah ZCT, Utah MGQT) and multiple issue (Utah MGQT): SR – if grand total ≤ -6 and all subtotals are negative; or any subtotal ≤ -3 NSR – when grand total ≥ +6 and all subtotals are positive INC – other results. 	<ul style="list-style-type: none"> • ZCT DI – when grand total ≤ -4 or if any subtotal ≤ -7 NDI – if grand total ≥ +2. Exception: consider as inconclusive if within test point difference of more than 7 points (e.g. R1:-2, R2:+6) INC – other results. • You-Phase (Bi-Zone): DI – if grand total ≤ -4 or sub-total ≤ -6 NDI – if grand total ≥ +4. Exception: consider as inconclusive if within test point difference of 7 points or more (e.g. R1:-3, R2:+6) INC – other results. • MGQT and DLST SR – if any subtotal ≤ -3 NSR – when all subtotals ≥ +1. INC – other results.
------------------------------	--	---	---

Table 3. Numerical polygraph charts evaluation and decision rules according to major TDA models: US Federal Government, University of Utah and Empirical Scoring System.

References

American Polygraph Association (2011). *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*: The Ad-Hoc Committee on Validated Techniques, Polygraph vol. 40, nr 4.

American Polygraph Association (2012). *Frequently Asked Questions about Validated Techniques and the Meta-Analysis* [online]. Available from: http://www.polygraph.org/files/validated_techniques_faq_1-9-2012.pdf [Accessed 16 November 2012].

Krapohl D., Handler M., Sturm S. (2012). *Terminology Reference for the Science of Psychophysiological Detection of Deception*, American Polygraph Association.

Krapohl D., McCloughan J., Senter S. (2006). *How to Use the Concealed Information Test*, *Polygraph*, vol. 35, nr 3.

National Research Council (2003). *The Polygraph and Lie Detection*, National Academies Press, Washington, DC.

Nelson R., Handler M., Shaw P., Gougler M., Blalock B., Russell Ch., Cushman B., Oelrich M. (2011). *Using the Empirical Scoring System*, *Polygraph*, vol. 40, nr 2, p. 67-78.

Weaver R.S. (1980). *The Numerical Evaluation of Polygraph Charts: Evolution and Comparison of Three Major Systems*, *Polygraph*, vol. 9, nr 2.