Laboratory Study of a Diagnostic Polygraph Technique in a Single Sequence: a replication study

Rodolfo Prado

Carlos Grajales

Raymond Nelson

Abstract

A previous laboratory study of a diagnostic polygraph technique in a single sequence reported results consistent with other validated techniques. This replication of our previous research included 163 new examinees and tested the effectiveness of an experimental single-sequence event-specific diagnostic polygraph technique with two relevant questions and evaluated with the Empirical Scoring System (Nelson et al., 2011). This replication experimental protocol had an unweighted accuracy of 86.1%, an 11.1% inconclusive rate, a sensitivity of 83.5%, a specificity of 88.7%. Reliability via Kappa's statistic was 0.722. The distributions of truthful and deceptive scores were not significantly different between the two studies. These findings are consistent with previous studies of event-specific diagnostic polygraph techniques with two relevant questions. Results add further support to the effectiveness of polygraph formats conducted in Spanish like those conducted in English.

Keywords: *diagnostic exams, directed-lie comparison question, lie detection, polygraph*

The previous project involved the study of decision accuracy of an experimental single-sequence diagnostic polygraph technique (Prado, Grajales, & Nelson, 2015). Accuracy of the experimental technique was 87%. Inspection of the confidence intervals reported herein and by APA (2011) indicated that the observed accuracy was consistent with the previously reported normal range of accuracy for diagnostic technique.

One of the limitations of that study was the fact that some of the examiners had only very recently completed their academic polygraph training and had virtually no actual field experience. Also, the study presented a large number of protocol violations resulting in unusable examination data. Nearly 20% of the examinations conducted could not be used due to heavily artifacted data that could not be interpreted and due to protocol violations on the part of the examiners. We attribute this to general inexperience on the part of many of the examiners and to the unfamiliarity of the examiners with an experimental test protocol for which the examiners had not received previous instruction or practice until the onset of this project.

The experimental format in one sequence did not outperform existing polygraph diagnostic format in any way and a confirmation of this false hypothesis is needed with a replication study. Another interesting and important aspect of the previous study was that all participants and examiners – and the first and second authors – were native Spanish speaking persons. Also, all of the examinations were conducted in Spanish, in Mexico City. This study is a replication of the earlier study with Spanish speaking persons in a different environment (in this case in Honduras).

Finally, although the previous study did not address the effectiveness of Directed Lie Comparison (DLC) questions, it is noteworthy that the study added support for the assumption that accuracy and effectiveness of Comparison Question Test (CQT) polygraph techniques using DLC questions can remain stable across language and cultural differences. Results of other studies of this experimental format, were needed to reach any conclusions about dimensions of criterion accuracy for this technique.

Methods

The present research project was designed to estimate the diagnostic accuracy this specific technique has in an analog laboratory setting. The study was performed between August 15th and August 21, 2013 using a convenience sample population. Half of the volunteers were cadets undergoing training at the Military Academy in Honduras and the other half were cadets training at the Police Academy of Honduras.

All the volunteers of the study were in between 18 and 23 years old and they all had 11 to 12 years of education. The volunteers were at the 2nd and 3rd year at the academy and almost 95% of them were males.

They all received an open invitation to participate in the study. The participants were told that:

- They could withdraw from the study at any time without punishment.

- They could contact the researchers for assistance if they experienced emotional discomfort from the study.

- They should inform any future polygraph examiner regarding their participation in this research project and they can advise others to contact the researchers if clarification is needed.

Interested participants were taken from the Military academy to the place where the study was conducted. Participants were required to:

1. be of legal age (18).
2. not to be under the influence of alcohol or drugs.
3. not to be excessively tired at the time of the test and.
4. not to be suffering from hunger at the time of the test.

Prior to volunteering, all subjects received a consent sheet, informing them of the use of an experimental polygraph format and the requirements of their participation in the activities. A total number of 163 volunteers participated in the study.

The subjects were randomly assigned a "status" by the Monitor of the study. Of the 163 original volunteers, 81 were assigned as "innocent" and 81 as "guilty". One case violated the protocol, and was then disregarded from the study.

The "guilty" status received instructions in Spanish that are summarized as follows:

> *You have been chosen as a "guilty" subject. You can decide if you don't want to continue in this study, but if you will, go inside that room and take a knife that you will find on a table there. In the same room you will find "Chavita". Chavita is the doll that you have to stab in repeated occasions until you damage her. Stab "Chavita as many times as you want, and make sure you do it well.*
>
> *After that, you will be taken to the polygraph room. The polygraph examiner will tell you that the test is about stabbing "Chavita". You have to deny any involvement in that activity during the duration of the polygraph evaluation. You have to lie when you say that you have nothing to do with that act.*
>
> *If the Polygraph evaluation can´t tell that you are guilty, and you obtain a "not deceptive" result, your reward will be one complete weekend out of the Academy. If the Polygraph evaluation tells that you are guilty, and you obtain a "deceptive" result, your punishment will be 15 days of arrest (not having free days out of the academy).*

The "innocent" status received instructions that are summarized as follows:

> *You have been chosen as an "innocent" subject. You can decide if you don't want to continue in this study, but if you will, go sit outside this room.*
>
> *You will be taken to the polygraph room. The polygraph examiner will tell you that the test is about stabbing "Chavita". You have to deny any involvement on that activity during the duration of the polygraph evaluation. You have to tell the truth when you say that you have nothing to do with that act.*
>
> *If the Polygraph evaluation tells that you are innocent, and you obtain a "not deceptive" result, your reward will*

be one complete weekend out of the Academy. If the Polygraph evaluation can´t tell that you are innocent, and you obtain a "deceptive" result, your punishment will be 15 days of arrest (not having free days out of the academy).

Both groups also received a final instruction in which it was made very clear that the reward or the punishment would be announced at the end of the test. Finally, all subjects were reminded that during the polygraph examination they were to deny their involvement in the false crime.

For the polygraph in-test, a Lafayette model LX4000 polygraph was used to record electrodermal activity (EDA), breathing movement, cardiovascular activity and voluntary activity. EDA was measured using skin resistance measured by standard Lafayette metal electrodes attached to the medial phalanges of the first and second fingers. Thoracic and abdominal breathing movement was measured by a standard Lafayette pneumatic tube assembly. Cardiovascular responses were recorded through the use of a Lafayette blood pressure cuff set at 80 to 90 mm of pressure and placed on the subject´s calf. A movement activity sensor pad was placed on the subject's seat.

The study was conducted at two separate facilities. Facility "A" was where the mock stabbing occurred. This was where the Monitor received the volunteers and assigned them their status. Once the subjects completed their field assignment, the "examiners coordinator" escorted them to the examiner who would conduct the test.

There were 25 independent examiners. The examiners' coordinator (coordinator) had the examiners on a list from 1 to 25 in facility "B" where the examiner had a communal working area and 6 polygraph evaluation rooms. The coordinator assigned the examinees to an examiner in order of appearance, and also assigned them a polygraph room. At the end of the test the coordinator accompanied the examinee to the waiting room in facility "A". Each exam was evaluated by both the examiner and the quality control reviewer. Quality Control examinations would result in a test being considered Non-Valid due to protocol violations that included:

- Physical illness or affliction in the examinee.

- Guilty subjects not appropriately denying or confessing to the crime.

- Examiners not correctly following the DLDT format.

- Interrupted or incomplete tests.

- Non-interpretable results.

Eleven examiners were Polygraph examiners with 1 year of experience and 14 examiners were students who had just graduated from their 10 week basic training program in polygraph examination. The quality control reviewer was the instructor leading the class during the last week of the training. An independent examiner with 20 years of experience conducted the "blind" Test Data Analysis of the charts.

After the tests and the quality control reviews were finished, the results (NDI, DI, INC, protocol violation) of the test were provided to the Monitor with the ground truth status inside an envelope with the name of the examinee. The coordinator sent that envelope back to facility "A", where the Monitor opened the envelope and compared the test result with the ground truth status. Examinees whose test results were truthful were rewarded with weekend time away from the academy regardless of their guilt status. No arrest or restriction consequences where actually imposed on examinees who did not produce truthful test results, though they were informed of the potential for restriction prior to their participation in the study activities.

Experimental Format

The test format included:

- Two neutral questions in positions 2 and 8 and repeated at positions 13 and 18

- One sacrifice relevant question in position 3

- Two relevant questions in positions 5 and 7 (first presentation), repeated at positions 10 and 12 (second presentation), 15 and 17 (third presentation), and finally in positions 20 and 22 (fourth presentation).

- Three DLC questions in positions 4, 6 and 9 (first

presentation), repeated at 11, 14 and 16 (second presentation), and finally at positions 19, 21 and 23 (third presentation).

The test questions, order and its type are summarized in Table 1.

Table 1. Questions presented (translated into English).

#	ID	Type	Text	Answer
1	X		The test is about to begin, please do not move and answer with yes or no to each question	
2	1N1	N	Are we in the year 2013?	Yes
3	SR	SR	Regarding the damage induced to Baby Chavita, do you intend to answer truthfully each question about that?	Yes
4	1C1	C	Have you ever hurt to a loved one?	No
5	1R1	R	Today, did you stab Baby Chavita?	No
6	1C2	C	Have you ever done something you later regretted?	No
7	1R2	R	Today, you stabbed Baby Chavita?	No
8	1N2	N	Are we in Tegucigalpa City?	Yes
9	1C3	C	Have you ever been irresponsible with your duties?	No
10	2R1	R	Today, did you stab Baby Chavita?	No
11	2C1	C	Have you ever hurt a loved one?	No
12	2R2	R	Today, you stabbed Baby Chavita?	No
13	2N1	N	Are we in the year 2013?	Yes
14	2C2	C	Have you ever done something you regretted later?	No
15	3R1	R	Today, did you stab Baby Chavita?	No
16	2C3	C	Have you ever been irresponsible with your duties?	No
17	3R2	R	Today, you stabbed Baby Chavita?	No
18	2N2	N	Are we in Tegucigalpa City?	Yes
19	3C1	C	Have you ever hurt a loved one?	No
20	4R1	R	Today, did you stab Baby Chavita?	No
21	3C2	C	Have you ever done something you regretted later?	No
22	4R2	R	Today, you stabbed Baby Chavita?	No
23	3C3	C	Have you ever been irresponsible with your duties?	No
24	XX		The test is about to end, please don't move until I release the air in the cuff	

Results

Inconclusive results are shown in Table 3, along with the 95% confidence intervals. Confidence intervals were obtained through two different approaches. The first is the computationally simple approach following Wilson, using a refinement of the simple asymptotic method. For the scoring, relevant questions were compared always against the comparison questions given *immediately before* the crime relevant question, using the ESS transformations, Two-Stage Decision Rules and cut scores for a 2 relevant question single issue test (Nelson et al., 2011).

From the 163 subjects that were first included in the study, only one case resulted in some form of protocol violation and was therefore excluded from the study calculations. The excluded protocol violation case had no significant impact on the sample size. The sample size that was subject to analysis after the exclusions is summarized in Table 2.

Table 2. Effective sample size in the study

SAMPLE SIZE	
Effective Sample Size	162
Subjects assigned to the "Innocent" Status	81
Subjects assigned to the "Guilty" Status	81

Table 3 Inconclusive results and estimated confidence intervals.

INCONCLUSIVE RATE	
Number of Inconclusive Results	18
Number of Inconclusive Results (Within "Innocent" Subjects)	10
Number of Inconclusive Results (Within "Guilty" Subjects)	8
Total Inconclusive Rate (Wilson's Confidence Interval)	11.111 % (7.145%, 16.879%)
Inconclusive Rate (Within "Innocent" Subjects) (Wilson's Confidence Interval)	12.345 % (6.846%, 21.255%)
Inconclusive Rate (Within "Guilty" Subjects) (Wilson's Confidence Interval)	9.876 % (5.090%, 18.296%)

Diagnostic accuracy was calculated excluding all inconclusive cases resulting in a sample size of 144 with, 71 programmed as "Innocent" and 74 programmed as "Guilty". Diagnostic Accuracy Measures obtained by the polygraph test are shown in Table 4.

Table 4. Accuracy profile.

DIAGNOSTIC ACCURACY	
Accuracy (Wilson's Confidence Interval	86.111 % (79.520%, 90.826%)
Sensitivity (Wilson's Confidence Interval)	83.561 % (73.429%, 90.339%)
Specificity (Wilson's Confidence Interval)	88.732 % (79.310%, 94.179%)
Error Rate (Wilson's Confidence Interval)	13.889 % (9.174%, 20.480%)
False Positives (Wilson's Confidence Interval)	5.555 % (2.842%, 10.579%)
False Negatives (Wilson's Confidence Interval)	8.333 % (4.831%, 14.001%)
Likelihood Ratio (+) (Confidence Intervals based on Risk Ratios)	7.42 (3.83 , 14.4)
Likelihood Ratio (-) (Confidence Intervals based on Risk Ratios)	0.185 (0.11 , 0.313)

This experimental format presented an accuracy profile that is comparable to those reported on the meta-analytic review (APA, 2011). Results show a respectable level of precision in the test, with accuracy results comparable and sometime exceeding those of other polygraph techniques. Estimates of diagnostic reliability obtained with the test are shown in Table 5.

Table 5. Diagnostic Reliability of the experimental format

DIAGNOSTIC RELIABILITY	
Kappa Statistic (Analytic Method Confidence Interval)	0.722 (0.610 , 0.835)
Area Under ROC Curve (Analytic Method Confidence Interval)	0.861 (0.805 , 0.918)
Agreement	86.11%
Correlation	0.7235

To provide methods for repeatability, a cross-tabulation of the test results is shown in Table 6. These numbers correspond to the basis of accuracy and reliability calculations, since inconclusive results are already excluded.

Table 6. Cross-tabulation of classification performance, excluding Inconclusive cases.

		PREDICTED CLASSIFICATION		
		Guilty	Innocent	TOTAL
STATUS	Guilty	61	12	73
	Innocent	8	63	71
TOTAL		69	75	TOTAL= 144 CASES

Results from the blind analysis of the data are shown in Table 7.

Table 7. Descriptive Statistics of Calculated Scores

CALCULATED SCORES	
Arithmetic Mean of Scores	0.6234
Standard Deviation of Scores	7.279
Arithmetic Mean of Scores (Within Innocents)	5.135
Standard Deviation of Scores (Within Innocents)	5.442
Arithmetic Mean of Scores (Within Guilty)	-3.889
Standard Deviation of Scores (Within Guilty)	5.983

Comparison with other diagnostic Techniques

Finally, the estimated profile accuracy of the experimental format presented in this research was compared with the mean results reported in the meta-analytic review (APA, 2011) for diagnostic techniques, excluding outliers. This comparison is shown in Table 8.

Table 8. Comparison of Accuracy Profiles among different techniques.

TECHNIQUE				
Diagnostic Acuracy Criterion	DLDT/ESS	Federal You-Phase/ ESS	Utah PLT (Combined)/ UTAH	ZCT/ESS
Acuracy	86.111 %	90.4%	93.0%	92.1%
Sensitivity	83.561 %	84.5%	85.3%	81.7%
Specificity	88.732 %	75.7%	80.9%	84.6%

By using exact binomial tests it was possible to verify statistically significant differences between the diagnostic measures obtained with the experimental format and other similar approaches already included in the meta-analysis review.

The experimental format presented no statistical difference in its estimated accuracy to that estimated by the Federal You-Phase technique (test's p-value=0.088). There is evidence of a slightly significant difference in accuracy between the experimental format and the ZCT/ESS technique (test's p-value=0.012). Also there is a highly significant difference with the accuracy of the experimental format technique (test's p-value=0.002) and the Utah.

For sensitivity, there are not statistically significant differences concluding that there is no evidence to assume that experimental format technique's sensitivity is lower than for any of the other techniques (all test's p-values>0.60).

The experimental format specificity results were significantly higher than the Federal You-Phase technique (test's p-value=0.008). There was no statistically significant difference compared with the Utah PLT or the ZCT/ESS techniques (test's p-value=0.098 and p-value=0.411, respectively).

Based on these three diagnostic accuracy criteria, there is no evidence to suggest that the experimental format technique has different accuracy than that of the Federal You-Phase. It is not different from the ZCT/ESS, nor Utah PLT techniques in terms of sensitivity and specificity. The only compared techniques that provided evidence of statistically better results was the Utah PLT (Combined) and the ZCT/ESS techniques, only due to a higher level of accuracy, since the sensitivity and specificity are no different than those of the experimental format. This seems to indicate that both techniques may only yield a marginal improvement over experimental format.

Table 9. Criterion Accuracy Profile

Criterion Accuracy Profile	
N Deceptive	73
N Truthful	71
Total N	144
Number Scorers	1
N of Deceptive Scores	69
N of Truthful Scores	75
Total Scores	144
Mean D	-3.889
Std Dev D	5.983
Mean T	5.135
Std Dev T	5.442
Reliability – Kappa	0.722
Reliability – Agreement	0.861
Reliability – Correlation	0.723
Unweighted Average Accuracy	0.861
Unweighted Average Inconclusives	0.111
Sensitivity	0.835
Specificity	0.887
FN Errors	0.083
FP Errors	0.055
D INC	0.098
T INC	0.123
Likelihood Ratio (+)	7.42
Likelihood Ratio (-)	0.185
D CORRECT	0.8356
T CORRECT	0.8873

Comparison of results between different number of presentations

We conducted a further analysis to complement the results of the study. This section presents the results of a series of statistical comparisons investigating differences the number of stimulus presentations may have had in the test results.

The first comparison was to test the impact that the number of presentations had in the inconclusive rates. These results are presented in Table 11. It is worth remembering that, after excluding the invalid case, 162 subjects were included in the sample, with 81 of these belonging to the "Innocent" group and 81 belonging to the "Guilty" group. See table 10.

Table 10. Comparison of the Effective sample size in the study.

SAMPLE SIZE	
Effective Sample Size	161
Subjects with the "Innocent" Status	81
Subjects with the "Guilty" Status	80

The first comparison was to test the impact that the number of presentations had in the inconclusive rates. These results are presented in Table 11. It is worth remembering that, after excluding the invalid case, 162 subjects were included in the sample, with 81 of these belonging to the "Innocent" group and 81 belonging to the "Guilty" group.

Table 11. Comparison of the Inconclusive rates between two levels of presentations three presentations (PRES) versus the use of three and four presentations of the test questions.

INCONCLUSIVE RATE		
	3 PRES	Using up to 4 PRES
Number of Inconclusive Results	58	18
Number of Inconclusive Results (Within "Innocent" Subjects)	31	10
Number of Inconclusive Results (Within "Guilty" Subjects)	27	8
Total Inconclusive Rate	35.80 %	11.11 %
Inconclusive Rate (Within "Innocent" Subjects)	38.27%	12.34%
Inconclusive Rate (Within "Guilty" Subjects)	33.33 %	9.88 %

According to the results in the table above, along with the results of a probability test on the equality of proportions using a large-sample statistic, there is a statistically significant decrease in the number of inconclusive results when using up to four presentations, compared to the numbers obtained with only three presentations (test's p-value<0.0001). The difference in the inconclusive rate is evident on both Innocent and Guilty Subjects (for both cases, tests' p-value ≈ 0.0001)

The results in Table 12 aim to verify whether the use of different numbers of presentations affect the diagnostic accuracy of the test in any way. The numbers are shown below.

Table 12. Comparison of the Diagnostic Accuracy of the test.

DIAGNOSTIC ACCURACY		
	3 Presentations	4 Presentations
Accuracy	85.43 %	85.71 %
Sensitivity	92%	83.6%
Specificity	79.24 %	87.7%
Error Rate	14.56%	14.28%
False Positives	10.67%	6.34%
False Negatives	3.88%	7.93%
Likelihood Ratio (+)	4.43	6.79
Likelihood Ratio (-)	0.101	0.187

According to the results, there is no statistical difference between the numbers obtained with three presentations and with four presentations (all test's p-values>0.10), in any of the accuracy measures presented in the table above. This evidence indicates that the difference is either negligible or too small to be detected by our experiment.

Distribution of the scores

Finally, the results in Table 13 indicates that the distributions of truthful and deceptive scores were not significantly different between this replication study and the previous one (Prado, Grajales, & Nelson, 2015).

Table 13. Distributions of truthful and deceptive scores

DIAGNOSTIC RELIABILITY	Experiment One	Experiment Two
Arithmetic Mean of Scores	1.38	0.6234
Standard Deviation of Scores	6.934	7.279
Arithmetic Mean of Scores (Within Innocents)	5.449	5.135
Standard Deviation of Scores (Within Innocents)	5.545	5.442
Arithmetic Mean of Scores (Within Guilty)	-3.256	-3.889
Standard Deviation of Scores (Within Guilty)	5.35	5.983

Conclusions

This replication study provides additional evidence that a single sequence technique did not out-perform traditional diagnostic techniques. Hypothesized advantages of a single sequence diagnostic format, beginning with the potential for increased test effectiveness that may result from reducing a source of uncontrolled response variance when starting and stopping the recording when using traditional diagnostic CQT formats cannot be confirmed by these studies. There may be no real advantage of single recording polygraph formats compared with multiple chart formats.

Finally, although the previous study did not address the effectiveness of DLC questions, it is noteworthy that this study adds support for previous finding in which the accuracy and effectiveness of polygraph evaluations conducted in Spanish are similar as those conducted in English. The accuracy of the polygraph evaluation remain stable across language and cultural differences. This study provides further support that DLC questions are robust, even with "inexperienced" examiners. We found no difference of accuracy between examiners with or without experience.

Though it was not the goal of this project we placed the cardio cuff on the lower leg and we found that this generates similar results as evaluations for which the cuff was placed on the arm, though with a less discomfort experience for the examinee.

Finally, we found that with three presentations of each question, the experimental format generated 8% inconclusive results. With four presentations it was reduced to 4%. Most of the inconclusive results involved innocent examinees. The 4th presentation didn't significantly affect accuracy, sensitivity and specificity of the test.

This study was limited in scope, and intended only as an attempt to replicate the results of an earlier study using this experimental diagnostic polygraph format in which multiple presentations of the question stimuli are accomplished in a single recorded sequence. This study did not compare the effectiveness of DLC and PLC methods, and did not compare the effectiveness of arm cuff and leg cuff response data. This study also made no attempt to define or investigate the psychological or physiological basis of responses to polygraph stimuli and addressed only a limited range of research questions regarding the accuracy of categorical test results and mean scores. These limitations notwithstanding, we conclude that these study results provide further support for the effectiveness of the polygraph in general, and further support for the effectiveness of DLC polygraph formats with exams conducted with native Spanish speaking persons. Although there is no advantage to the use of the experimental format compared with other validated polygraph formats, we recommend continued research and continued interest in the potential for the development of a further improved single sequence single issue diagnostic polygraph format.

Funding for this study was provided by the IPTC.

We want to thank to the Consejo Nacional de Defensa y Seguridad de Honduras, and to the Polygraph Unit of the Dirección Nacional de Investigación e Inteligencia de Honduras without whose participation this study would not have been possible.

References

Abhyankar, Smruti & Kaur, Gursharn (2010). Confidence interval for Binomial Proportions. .Retrieved on January 5, 2015 at [http://www.isid.ac.in/~deepayan/SC2010/project-sub/bootstrap_binomial_report.pdf].

American Polygraph Association. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph, 40, 194-305.

American Polygraph Association (2009a). *Model Policy for Post-conviction Sex Offender Testing.* [Electronic version] Retrieved January 25, 2012, from http://www .polygraph.org

Ansley, N. (1998). The zone comparison test. Polygraph, 27, 108-122.

Barland, G. (1981). A validation and reliability study of counterintelligence screening tests. U.S. Department of Defense, Security Support Batallion. 902nd Military Intelligence Group, Ft. Mead, MD.

Bossuyt, Patrick et al. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ 2003;326:41.1.

Kircher, J. C., 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*, 22–67.

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.

Nelson, R. (2015). The scientific basis for polygraph testing. Polygraph, 44(1) 28-61.

Nelson, R., Handler, M., Adams, G., & Backster, C. (In press). Survey of reliability and criterion validity of Backster numerical scores of You-Phase exams from confirmed field investigations. Polygraph.

Nelson, R. Handler, M. Shaw, P., Gougler, M., Blalock, B., Russell, C., Cushman, B., & Oelrich, M. (2011). Using the Empirical Scoring System, *Polygraph*, 40 (2).

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

Prado, R., Grajales C., & Nelson R. (2015). Laboratory Study of Directed Lie Polygraphs with Spanish Speaking Examinees. Polygraph. 44(1), 79-90.

R Development Core Team, (2008). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.