Polygraph

VOLUME 44 2015 NUMBER 2

Contents

Published Semi-annually

P.O. Box 8037, Chattanooga, Tennessee 37414-0037

Polygraph

APA Officers for 2015-2016

Subscription information: *Polygraph* is published semi-annually by the American Polygraph Association. Editorial Address is Editor@polygraph.org. Subscription rates for 2015: One year $150.00 (Domestic), $180.00 (Foreign). Change of address: APA National Office, P.O. Box 8037 Chattanooga, TN 37414-0037. THE PUBLICATION OF AN ARTICLE IN *POLYGRAPH* DOES NOT CONSTITUTE AN OFFICIAL ENDORSEMENT BY THE AMERICAN POLYGRAPH ASSOCIATION.

Instructions to Authors

Scope

The journal *Polygraph* publishes articles about the psychophysiological detection of deception, and related areas. Authors are invited to submit manuscripts of original research, literature reviews, legal briefs, theoretical papers, instructional pieces, case histories, book reviews, short reports, and similar works. Special topics will be considered on an individual basis. A minimum standard for acceptance is that the paper be of general interest to practitioners, instructors and researchers of polygraphy. From time to time there will be a call for papers on specific topics.

Manuscript Submission

Manuscripts must be in English, and may be submitted, along with a cover letter, on electronic media (MS Word). The cover letter should include a telephone number, and e-mail address. All manuscripts will be subject to a formal peer-review. Authors may submit their manuscripts as an e-mail attachment with the cover letter included in the body of the e-mail to:

Editor@polygraph.org

As a condition of publication, authors agree that all text, figures, or other content in the submitted manuscript is correctly cited, and that the work, all or in part, is not under consideration for publication elsewhere. Authors also agree to give reasonable access to their data to APA members upon written request.

Manuscript Organization and Style

All manuscripts must be complete, balanced, and accurate. Authors should follow guidelines in the *Publications Manual of the American Psychological Association*. The manual can be found in most public and university libraries, or it can be ordered from: American Psychological Association Publications, 1200 17th Street, N.W., Washington, DC 20036, USA. Writers may exercise some freedom of style, but they will be held to a standard of clarity, organization, and accuracy. Authors are responsible for assuring their work includes correct citations. Consistent with the ethical standards of the discipline, the American Polygraph Association considers quotation of another's work without proper citation a grievous offense. The standard for nomenclature shall be the *Terminology Reference for the Science of Psycho-physiological Detection of Deception* (2012) which is available from the national office of the American Polygraph Association. Legal case citations should follow the *West* system.

Manuscript Review

An Associate Editor will handle papers, and the author may, at the discretion of the Associate Editor, communicate directly with him or her. For all submissions, every effort will be made to provide the author a review within 4 weeks of receipt of manuscript. Articles submitted for publication are evaluated according to several criteria including significance of the contribution to the polygraph field, clarity, accuracy, and consistency.

Copyright

Authors submitting a paper to the American Polygraph Association (APA) do so with the understanding that the copyright for the paper will be assigned to the American Polygraph Association if the paper is accepted for publication. The APA, however, will not put any limitation on the personal freedom of the author(s) to use material contained in the paper in other works, and request for republication will be granted if the senior author approves.

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.

A Literature Review of Polygraph Countermeasures and the Comparison Question Technique

Mark Handler and Charles Honts

Honts, Handler, & Hartwig, LLC.

Walt Goodson

Texas Department of Public Safety

Abstract

We reviewed the research of countermeasures effect on the comparison question technique. We provide a consolidation of countermeasure literature as well as an operational definition and taxonomy of countermeasures. We surveyed the pertinent literature regarding the effectiveness and limitations of certain countermeasure tactics. We offer evidence-based answers are to common countermeasures questions and make recommendations for reporting countermeasures.

Keywords: *countermeasures, polygraph comparison question technique*

A Literature Review of Polygraph Countermeasures and the Comparison Question Technique

In order for a countermeasure to be effective in a Comparison Question Technique (CQT), it must satisfy two requirements. First, it must create a sufficient difference in the polygraph measurements to comparison and relevant questions to produce a truthful or inconclusive outcome. Secondly, it must be done covertly as to not be identified by the examiner, an observer, or any quality control review. In considering what information would be most helpful to examiners we provide evidence-based answers to some important questions about countermeasures.

Our operational definition of "countermeasure"?

There have been a number of proposed definitions from within and outside of the profession for the term countermeasure (CM). We needed to operationally define CM as it applies to polygraph testing. For our purposes, we considered a CM to be anything a test subject does in an attempt to alter the test data so as to produce a truthful (negative) test result. This definition encompasses the truthful subjects trying to ensure a True Negative (TN) result and the deceptive subjects trying to produce a False Negative (FN) outcome. One could ostensibly argue that all subjects engage in some form of behavior to produce truthful outcomes and are thus attempting CMs - the truthful tell the truth and the deceptive lie, but we feel these actions don't fit our definition for altering the test data. To alter means *to change or make different* in a meaningful way.

What type of CMs do people use?

We followed Honts' (1987) taxonomy as it breaks down CMs into categories that have been researched, though others have produced different recommendations for CM categorization (see Krapohl, 2009). In following Honts (1987) we break CMs down into the following categories;

1. General State CMs- actions intended to alter the subject's psychological state and/or measured physi-

Authors' note: A portion of this document appeared previously in APA Magazine 48(2) under Handler, Honts & Blalock. The authors are grateful to Pam Shaw and Don Krapohl for their thoughtful comments, suggestions and critical reviews of earlier drafts of this manuscript. The thoughts, ideas, and recommendations do not necessarily reflect those of the American Polygraph Association or the Texas Department of Public Safety. Correspondence should be addressed to Mark Handler, APA editor in chief. Contact: editor@polygraph.org

ological responses throughout the entire examination. They include such things as; drugs, relaxation, or interfering agents. They are not focused on any specific point in the testing.

2. Specific Point CMs- as their name suggests, these are actions the subject takes at specific points in the testing process. They can be attempts to reduce responses to relevant questions but are usually efforts to increase responses to comparison questions. They can be employed physically, mentally or in combination.

3. Spontaneous CMs- these are CMs that subjects report doing without planning or forethought. A number of laboratory studies debriefed subjects about efforts to produce truthful outcomes. These debriefs are the source of most of our knowledge of spontaneous CMs. Subjects report trying such things as; relaxation, rationalization, imagery, attempts to control their breathing or heart rate, trying to stay calm, biting their tongue and pressing their toes at random places.

4. Information CMs- people who know they are going to take a polygraph examination (both guilty and innocent) often seek information about polygraph techniques and CMs from the internet or other sources. This information-seeking can be motivated by an attempt to satisfy curiosity, to try and hide deception, or in an effort to ensure that truthfulness is obvious.

Given our operational definition and taxonomy we sought to provide evidence-based answers to some important questions about CMs. Evidence-based answers and practices concerning CMs are not simple. They have to be based upon research and not on anecdote or dogma. Evidence-based answers and practices have to be qualified by the limitations of the research upon which they are based. Those qualifications depend on such things as whether the subjects were coached or if they received practice on an instrument. Who were the subjects? Did the examiner use some sort of activity sensor? The following is a summary of some findings from the peer-reviewed published studies we examined for this paper.

1. Rovner (1986) is a rewrite of his 1979 doctoral dissertation in which he trained subjects on the principles of CQT testing, including giving them pictorial examples of reactions. He called these the *Info group*. He also gave the *Info group* Specific Point CM training using a variety of physical and mental CMs known to produce reactions. He had a second group called the *Info + Practice group*. He gave them the same material but allowed them to practice their CMs on a polygraph before their real test. The accuracy of the results for the control and the *Info group* was about 88%. However, the *Info + Practice group* accuracy results were about 62%. He did not report using an activity sensor or making attempts to identify CM subjects.

2. Dawson (1980) used Stanislavsky trained actors to attempt General State CMs in a mock-crime lab study. They were not trained in polygraph principles and they did not receive practice. The CM group actors were motivated to appear innocent to display their superior acting skills. CM deceptive subjects said they used imagery and prior memories as strategies. The General State CM effects were ineffective. Excluding inconclusive results, all CM subjects were found deceptive.

A side note, Dawson conducted an interesting additional experiment. He had subjects answer the polygraph questions in two ways, immediately after the question and delayed by 8 seconds. He measured responses in three ways; after the question without an answer, after the immediate answer and after the delayed answer. Based on numerical scores he reported the following. Immediate answers accuracy was 75%correct, 12% incorrect, and 12% inconclusive. The mea-

surements following the question but before the delayed answer produced 83% correct, 8% incorrect, and 8% inconclusive. The measurements taken after the delayed answer resulted in 29% correct, 8% incorrect, and 62% inconclusive.

3. Bradley & Ainsworth (1984) tested General State CMs by using alcohol intoxication during a crime act and also during polygraph testing. They tested subjects with the CQT and the CIT. We limit our discussion to the CQT findings. They measured heartrate, respiration and electrodermal responses during a mock-crime robbery and shooting. They reported alcohol intoxication during the crime decreased detectability with electrodermal activity. Intoxication during the testing was ineffective.

4. Honts, Hodes & Raskin (1985) in experiment 1 trained the CM group on the principles of polygraph CQT and Specific Point CMs. They coached the subjects on physical (press toes to floor) and pain (bite tongue) CMs. They did not get any practice on an instrument. They used a photoelectric plethysmograph instead of a cardio cuff. They reported no significant effects for the CM group. They were unable to detect which subjects used CMs by either direct observation or reviewing the charts. They did not use an activity sensor.

5. Honts, Hodes & Raskin (1985) in experiment 2 trained the CM group on the principles of polygraph CQT and Specific Point CMs. There they coached the subjects on physical (press toes to floor) and pain (bite tongue) CMs. However, unlike experiment 1, these CM subjects got to practice on an instrument. They used a standard cardio cuff in this experiment. They reported there was a 47% FN rate for the CM group. Again they were unable to detect which subjects used CMs by either direct observation or reviewing the charts. They did not use an activity sensor.

6. Honts, Raskin, & Kircher (1987) trained subjects in physical (press toes to floor) and pain (bite tongue) Specific Point CMs. They gave the subjects training on CQT principles and coaching on when and how to apply the CMs to the comparison questions. They did not give any practice on an instrument. They measured muscle movement by electromyography (EMG) on the subjects' jaw and calf. They reported no FNs with the guilty control group who did not use any CMs. They identified 78% of the truthful subjects correctly. Seventy percent of the guilty CM group produced FN results. They identified 90% of the CM subjects by EMG tracings.

7. Honts, Raskin, Kircher & Hodes (1988) included 65 deceptive subjects from four studies who were debriefed about any use of Spontaneous CMs. Sixty percent (39/65) of the deceptive subjects admitted to Spontaneous CMs. The strategies included: relaxation, rationalization, self-deception, disassociation, imagery, attempts to control breathing or heartrate, biting tongue, attempts to control general physiological responses and pressing toes to the floor. Blind scoring accuracy was 80% correct, 3% wrong and 17% inconclusive. The use of Spontaneous CMs did not affect the test results. Examiners were unable to differentiate CM users. No activity sensor was used.

8. Raskin and Kircher (1990) trained subjects in physical (muscle contraction) and mental (counting backwards) Specific Point CMs. They taught them the principles of polygraph testing and when they should employ the CMs. They coached them and they had prac-

tice on an instrument. They used a seat activity sensor. The CMs produced about 50% FNs when scored by the computer. All of the physical CMs were identified by reviewing the seat sensor data.

They also trained a "relaxation" group on the principles of polygraph. This group used autogenic relaxation as a General State CM throughout the entire examination. The relaxation General State CMs were not effective.

9. Honts, Raskin & Kircher (1994) trained subjects in Specific Point CMs and CQT polygraph principles. Here they used physical (muscle contraction), pain (bite tongue) and mental (counting backwards) CMs during the comparison questions. The subjects were coached but not given any practice on the polygraph instrument. They measured EMG on the jaw and calf. The CM group produced a 50% FN rate and were not detected by direct observation or by looking at the polygraph charts. They were able to identify the pain and physical CM groups by EMG scores which were significantly larger than the other groups.

10. O'Toole et al., (1994) was a partial replication of the Bradley & Ainsworth (1984) study on General State CMs using alcohol. Here the deceptive CM group were intoxicated only during the mock-crime theft. They measured skin conductance, respiration and finger pulse amplitude. They did not provide any information on polygraph principles, they did not coach or practice any CMs with the CM subjects. Alcohol intoxication during the crime had no effect on polygraph results.

11. Honts, Amato and Gordon (2001) did not train any of their subjects but debriefed them on what actions they took to help pass their polygraph tests in a large lab study. Overall 68% of the subjects reported they attempted a Spontaneous CM. Almost half of the truthful subjects (46%) reported using at least one Spontaneous CM. These Spontaneous CMs included altered breathing, mental and physical strategies. The Spontaneous CMs did not affect the deceptive scores but it shifted the truthful scores in a negative direction. In other words, the truthful subjects who tried to help increase their chance of a TN result had less truthful scores than those who did not attempt Spontaneous CMs. These subjects were not coached, nor were they given practice tests on an instrument. There was no activity sensor used in any of these cases.

    Later, three federally certified instructors reviewed the charts in an effort to identify the presence of CMs. None of the three federally trained instructors could identify the CM subjects at better than chance levels by reviewing the test data.

12. Oglivie & Dutton (2008) reanalyzed unpublished data from a CM lab experiment with and without the activity sensor tracing. In this case the activity sensors included, seat, feet and arm sensors. The CM polygraph subjects in these cases received a copy of the internet publication of *The Lie Behind the Lie Detector* (Maschke & Scalabrini, 2002) as a reference. This material has detailed descriptions of CQT polygraph principles and CM strategies. The CM subjects had to pass a written examination on the CM material before proceeding. They used physical (toe curling) Specific Point CMs during the presentation of the comparison questions. A subset of the CM group was allowed to practice on an instrument while receiving real-time feedback from an experienced examiner.

    Five experienced polygraph examiners reviewed the polygraph charts in two conditions; with and without the activity sensor data. The examiners were asked; a) to score the test data using a 7-po-

sition federal scoring rules, and b) using a 5-point Likert scale rate the absence or presence of CMs.

Mean CM present/absent scores were statistically significant between the sensor and no-sensor conditions. Without the activity sensor, the CM present/absent scores were near zero. Examiners did not perform above chance levels when relying on the traditional polygraph channel data to identify CMs without the activity sensor data. Adding the activity sensor data made identifying the CM group more likely.

Of note, the average guilty CM total scores with the activity sensor was similar to the truthful scores (they were positive). The average guilty CM total scores without the activity sensor was double the average truthful scores. This suggests that the Specific Point CM strategies can create increased FN results.

13. Honts & Reavy (2015) did not train any of their subjects but debriefed them on what actions they took to help pass their polygraph tests in a large lab study. Overall 48% of the subjects reported they attempted a Spontaneous CM. About 50% of the subjects who received probable lie comparison questions (PLC) reported Spontaneous CMs. For the directed lie comparison question variant (DLC) subjects, about 46% reported Spontaneous CMs. Breaking it down by guilt, overall 78% of the deceptive group reported Spontaneous CM attempts (DLC=72%, PLC=83%). Within the truthful group 18% overall reported attempts at Spontaneous CMs (DLC=20%, PLC=15%). None of the differences were significant.

    The investigators did not make any attempt to identify CMs either through direct observation or data review. Just like in Honts, Amato, & Gordon (2001) the scores for the truthful CM groups were less positive. Once again this suggests truthful subjects who use Spontaneous CMs hurt their chances of passing the test.

14. Stephenson & Barry (1998) was not a CQT study but is discussed here due to the findings. The subjects were polygraph examiners so they had extensive knowledge about the CQT polygraph procedures and CMs. They tested whether they could identify when a subject makes a "physical movement between the toes and shoulder" while collecting polygraph test data. They used an activity sensor mounted to the front of the test subject's chair. They collected test data in a mock acquaintance test during which the examiner counted from 1 through 10. At some point the test subject was to make a covert physical CM attempt and remember the number at which they attempted the CM. They stationed an "observer" directly in front of the test subject who watched for movement. The overall CM detection rate for the examiner observing the subject was 9%. The "observer" identified the Specific Point CMs in 36% in of the cases. When they used the movement sensor tracing, the experimenters identified 85% of the CM attempts.

15. Honts & Alloway (2007) was a constructive replication of Rovner (1986) using the Test for Espionage and Sabotage (TES). They gave half of the truthful and deceptive subjects the book *The Lie Behind the Lie Detector* (Maschke & Scalabrini, 2000) to study for one week. They found no significant effect of providing the Information CM material on validity of the TES. Once again, however, those subjects who reported using CMs had significantly lower probabilities of truthful scores. This included both the truthful and the deceptive subjects.

Here we will provide some common CM questions along with evidence-based answers.

1. *Do both truthful and deceptive subjects attempt CMs?*

The simple answer is "Yes." Research shows that both truthful and deceptive subjects report attempting Spontaneous CMs. From study 9 (Honts, Raskin & Kircher 1994) which was limited to deceptive subjects, 65% of them attempted Spontaneous CMs. From study 13 (Honts & Reavy, 2015) about half of the subjects overall reported attempting Spontaneous CMs. A larger proportion of deceptive subjects reported attempting Spontaneous CMs but 18% of truthful subjects also reported attempting Spontaneous CMs. From study 11 (Honts, Amato and Gordon, 2001) we see about 68% overall and about 50% of truthful subjects attempted Spontaneous CMs.

2. *What type of CMs do subjects attempt?*

From a number of studies above Spontaneous CMs include a variety of reported strategies; relaxation, rationalization, self-deception, disassociation, imagery, attempts to control breathing or heartrate, biting tongue, attempts to control general physiological responses and pressing toes to the floor. Specific Point CMs generally included physical (press toes, curl toes, etc.) or pain (biting tongue) and mental (counting backwards) activities. Some Information CM sources suggest such actions as squeezing the anal sphincter (http://www.polygraph.com/). More sophisticated advice about examination behavior and chart recording CMs is offered at https://antipolygraph.org/ (Maschke & Scalabrini, 2005). Some examinees reported attempting a form of General State CMs when they describe attempts at rationalization, relaxation, disassociation, imagery, etc.

3. *What type of CMs are effective at increasing TN results, creating a FN result, or resulting in an inconclusive outcomes- and to what degree?*

Spontaneous CM produced no effects for the deceptive subjects in terms of increased TN or inconclusive outcomes, nor were there reliable effects found in the numerical scores. Deceptive subjects in study 15 shifted the scores away from a truthful result. Spontaneous CMs by truthful subjects decreased their chances of being found truthful. Information CMs that lead to Spontaneous CMs simply shifted truthful scores in the negative direction (see study 15). General State CMs have not been shown to be effective, see study 2 and 10. Study 3 reported some effect for intoxication during the mock crime act. Specific Point CMs have been shown to be effective in shifting differential response measurements and increasing FN results (see studies 1, 4, 5, 6, 8, 9, & 12) following specific training, but not just information. Specific Point CMs thus seem to be most dangerous when coupled with hands-on training and practice.

4. *Do polygraph test subjects attempt CMs more with Directed Lie Comparison questions versus the Probable Lie variant?*

This has not been shown by the relevant research (see study 13).

5. *Can examiners identify examinees using CMs at better than chance rates? And does the addition of activity sensors make a difference?*

Without an activity sensor there are no studies that support examiners can identify CMs at better than chance rates (see studies 4, 5, 7, 11, & 12). In fact, the research indicates that when examiners try to identify countermeasure they falsely accuse a substantial number (47% or more) of innocent non-countermeasure users of using CMs (study 5). With an activity sensor (or EMG) polygraph examiners are able to significantly identify CM users (see studies 6, 8, 9, 12, & 14) who use CMs that required movement (for example, pressing the toes to the floor.) Finally, there is no evidence that current training in countermeasure detection is effective. In fact the alleged respiratory countermeasure signatures caused by the countermeasure materials produced by Williams (http://www.polygraph.com/) have been shown to occur naturally in a substantial number of actually innocent subjects who were not using CMs (Honts & Crawford, 2010).

6. *How does using CMs affect the scores of truthful and deceptive subjects?*

Specific Point CMs increase FN outcomes following training by producing significant effects in all of the polygraph components depending upon the countermeasure used (see studies 1, 5, 6, 8, 9, & 12). It is unclear what their effect would be for increasing TN outcomes, though there is no reason to think they would not be effective.

Spontaneous CMs don't increase FN and probably decrease TN results. Information

CMs that lead to Spontaneous CMs would be expected to have similar results. Spontaneous CMs are extremely common with examinees and there does not appear to be any evidence that such CMs are effective. Therefore, as the evidence seems to suggest, if the data simply appears to be messy, and there is *sufficient* uncontaminated data to conduct an analysis, the scorer should attempt to analyze the uncontaminated data, and a decision should be rendered by the scorer if conclusive scores are reached (ex. NDI/NSR, DI/SR). Examiners should report when data quantity and quality are insufficient to complete a standardized numerical evaluation. An example of reporting language is:

> *After assessing the quantity and quality of the test data collected in this examination, I determined that the test data were of insufficient interpretable quantity and/or quality as a result of numerous artifacts to conduct a standard numerical evaluation. In other words, there was insufficient data to evaluate in order to render a reliable decision on this examination.*

General State CMs are unlikely to create a differential response between relevant and control questions that would increase TN or FN results. At worst they might be expected to cause an inconclusive result due to mitigating the overall responsivity to all test questions, but even increases in inconclusive outcomes have never been demonstrated in a published peer-reviewed study. An unpublished study (Gatchel et al., 1983) tested the General State CM effects of the beta-blocker drug propranolol. The only significant finding was an increase in accuracy with the innocent. Study 3 reported no effect for alcohol intoxication during a polygraph test. However, as mentioned, they reported an effect for intoxication at the time of the crime. The replication of that study failed to find an effect for alcohol and FN results for intoxication at the time of the crime (see study 10). In study 2 experienced actors try to produce FN results using General State CMs but produced no effect.

In summary the CM research base is incomplete and additional research is needed. However, the limited research shows trained CMs are something that should concern examiners as under certain circumstances they have produced substantial numbers of FN errors. Moreover when trained deceptive subjects use CMs, examiners have not shown an ability to identify those subjects at better than chance rates without some sort of activity sensor (and then only for CMs that require physical movement). Regardless of any alleged anecdotal successes at detecting CMs, no research has shown that any examiner can reliably detect CMs from simple pattern recognition. In fact, as mentioned, research has shown that the respiratory patterns that are allegedly linked to some internet training approaches occur naturally in the respiration recordings of a substantial number of actually innocent subjects (Honts & Crawford, 2010).

We realize a number of things that might be CMs appear spontaneously among truthful examinees. What may distinguish these events from CMs, though, is the frequency or the targeting of the behaviors. For example, both truthful and deceptive examinees move during polygraph tests. This does not, in and of itself, mean that movements are not useful in detecting CM attempts. Indeed, research shows that movements can be strong indicators in that regard. The mere presence of hyperventilation, as another example, does not confirm CMs, but if they persist despite examiner warnings or they seem to appear only on one category of question, then they can be useful indicators. Ultimately we hope further research will help develop improved objective measures of anomalies among groups of questions. Future CM detection efforts should probably seek such an objective measurement approach.

The research clearly shows that when examiners do try to detect CMs they falsely accuse a substantial number of actually innocent subjects. Examiners should be extremely cautious about reporting CMs based on their ability to intuit a subject has used CMs. Doing so puts the innocent at risk. The upside to this literature is that when deceptive subjects engage in CMs that require movement they can be reliably identified when examiners use an activity sensor. Finally, there is no published research that information provided by internet CM websites is at all dangerous to the validity of the CQT.

References

Bradley, M.T. & Ainsworth, D., (1984). Alcohol and the psychophysiological detection of deception. Psychophysiology 21, 63–71.

Dawson, M.E. (1980). Physiological detection of deception: Measurement of responses to questions and answers during countermeasure maneuvers. Psychophysiology 17, 8–17.

Gatchel, R. J., Smith, J. E., Kaplan, N. M., et al. (1983). The effect of propranolol on polygraphic detection of deception. Unpublished manuscript.

Honts, C.R. (1986). Countermeasures and the physiological detection of deception: a psychophysiological analysis. Dissertation Abstracts International 47, 1761B.

Honts, C.R., Hodes, R.L. & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. Journal of Applied Psychology 70, 177–187.

Honts, C.R. (1987). Interpreting research on countermeasures and the physiological detection of deception. Journal of Police Science and Administration 15, 204–209.

Honts, C.R., Raskin, D.C. & Kircher, J.C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. Journal of Psychophysiology 1, 241–247.

Honts, C.R., Raskin, D.C., Kircher, J.C. & Hodes, R.L. (1988). Effects of spontaneous countermeasures on the physiological detection of deception. Journal of Police Science and Administration 16, 91–94.

Honts, C.R., Raskin, D.C. & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. Journal of Applied Psychology 79, 252–259.

Honts, C.R., Amato, S. & Gordon, A.K. (2001). Effects of spontaneous countermeasures used against the comparison question test. Polygraph 30, 1–9.

Honts, C.R. & Alloway, W. (2007). Information does not affect the validity of a comparison question test. Legal and Criminological Psychology 12, 311–312.

Honts, C.R. & Crawford, M. (2010). Polygraph countermeasures cannot be detected from respiratory signatures: Government policy puts the innocent at risk. Paper presented at the American Psychology-Law Society Meeting, Vancouver. 17–20 March.

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

Krapohl, D.H., (2009). A taxonomy of polygraph countermeasures. Polygraph 38, 89–105.

Maschke, G. W. & Scalabrini, G. J. (2002). The lie behind the lie detector. Internet https://antipolygraph.org/lie-behind-the-lie-detector.pdf

Maschke, G. W. & Scalabrini, G. J. (2005). The lie behind the lie detector. Second Edition. Internet: https://antipolygraph.org/lie-behind-the-lie-detector.pdf

Ogilvie, J. & Dutton, D. W. (2008) Improving the detection of physical countermeasures with chair sensors. Polygraph 37(4), 136-148.

O'Toole, D., Yuille, J.C., Patrick, C.J. & Iacono, W.G. (1994). Alcohol and the physiological detection of deception: arousal and memory influences. Psychophysiology 31, 253–263.

Raskin, D. C. & Kircher J. C. (1990) Development of a computerized polygraph system and physiological measures for detection of deception and countermeasures: A pilot study.

Unpublished Manuscript.

Rovner, L.I., Raskin, D.C. & Kircher, J.C. (1979). Effects of information and practice on detection of deception. Psychophysiology 16, 197–198 (abstract).

Rovner, L.I. (1986). The accuracy of physiological detection of deception for subjects with prior knowledge. Polygraph 15, 1–39.

Stephenson, M., Barry, G. (1988). Use of a motion chair in the detection of physical countermeasures. Polygraph, 17(1), 21-27.

Examiners may find Table 1 a quick reference for a consolidation of the CM study data. Note Honts et al., 1988 is not included in the table as those results were derived from included studies.

Table 1 – Breakdown of CM study findings.

Study	Test type	Type of CM	Training Y/N	Coached/ Practice on inst. Y/N	Activity Sensor Y/N	Findings reported
Rovner (1986)	CQT	Practice CM group used physical and mental CMs. They got to practice and received feedback.	Info group and info + practice group All given extensive training on polygraph principles and CN strategies	Coached-yes Practice-yes	No	Accuracy of scoring: Standard group = 87.5%; Info group = 87.5%; Info+practice group = 62.5%
Dawson (1980)	CQT	General State CMs	No	No	No	General State CMs had no effect.
Bradley & Ainsworth (1984)	Limited to CQT part	General state CMs-alcohol intox during crime and during polygraph	No	No	No	No effect for intox during testing. EDA responses were reduced for intox during crime scenario.
Honts, Hodes & Raskin (1985) Experiment number 1	CQT	CM groups used: 1.Bite tongue or 2.muscle contraction press toes to floor	All given extensive training on polygraph principles and CN strategies	Coached-yes Practice-no	No Also no BP cuff used in experiment 1	No significant effects for CMs found. Unable to detect CM subjects through chart interpretation or observation.
Honts, Hodes & Raskin (1985) Experiment number 2	CQT	CM groups used: 1.Bite tongue or 2.muscle contraction press toes to floor	All given extensive training on polygraph principles and CN strategies	Coached – Yes Practice-Yes	No	47% FN for CM groups. Unable to detect CM subjects through chart interpretation or observation.
Honts, Raskin & Kircher (1987)	CQT	CM group used: bite tongue & press toes to floor	Yes CM group trained on polygraph principles and CM strategy	Coached-yes Practice-no	Yes EMG on the jaw and leg.	No FN with Guilty control. 78% innocent correct. 70% FNs. 90 % of CM subjects identified with the EMG tracings.
Raskin & Kircher (1990)	CQT	CM group used: Physical-muscle contraction PLUS Mental Counting backwards Relaxation group used: mental imagery	Yes CM group and relaxation group trained on polygraph principles and CMs. Practice recording, practice logs,	Coached-yes Practice-no	Yes seat sensor	Computer: CM group =50% FN. Relax not effective. All physical CM subjects identified by activity sensor.
Honts, Raskin & Kircher (1994)	CQT	CM group used: Physical-muscle contraction Pain bite tongue Mental group: Counting backwards	All given extensive training on polygraph principles and CN strategies	Coached-yes Practice-no	EMG-jaw and calf	FN= 50% for Mental & Phys. CM Unable to detect CMs by observation or chart tracing EMG measures for CM groups were significantly larger than controls.
O'Toole et al. (1994)	Limited to CQT part	General state CMs-alcohol intox during crime	No	No	No	No effect for intox during crime
Study	Test type	Type of CM	Training Y/N	Coached/ Practice on inst. Y/N	Activity Sensor Y/N	Findings reported

Study	Test type	Type of CM	Training Y/N	Coached/ Practice on inst. Y/N	Activity Sensor Y/N	Findings reported
Honts, Amato & Gordon (2001)	CQT	All CMs were spontaneous- Some subjects used more than one. 32%-breathing 76%-mental 10% physical	No	Coached-no Practice-no	No	3 DoDPI instructors did not identify presence of CMs above chance level. 47% of CM present decisions were on truthful subjects.
Oglivie & Dutton (2008	CQT	Physical CMs: Toe curling	All given extensive training on polygraph principles and CN strategies	Coached-yes Practice yes for a subset	Compared with and without seat, feet, and arm activity sensor data. 15 Innocent. & guilty control and 38 CMs. Charts printed with and without activity trace.	Five scorers looked at 68 cases Mean activity sensor scores significant for CM group with and without sensor. Unable to determine presence of CMs without the sensor data.
Honts & Reavy(2015)	CQT	spontaneous	no	no	yes	No attempt to identify CMs. 48% attempted CM PLC=50%; DLC=46% 78% Guilty attempted PLC 83%; DLC 72% 18% Innocent attempted PLC 15%; DLC20% Truthful scores less positive when CMs attempted
Stephenson & Barry (1988)	Examiner counted 1-10 and	Subject made a physical movement between the toes and shoulder	Info yes because it was an examiner.	Coached-yes because it was an examiner. Practice-yes	Yes-used Lafayette chair with activity sensor bar under the front legs	Had an observer in front of subject in addition to examiner. CM detection rates were: Examiner=9%; observer=36%; movement sensor tracing= 85%
Honts & Alloway (2007)	CQT	Information and Spontaneous CMs	No	No	Yes under the chair legs	No effect on FN Deceptive and Truthful CM subjects' probabilities scores moved away from truthfulness.

A book review of *Investigative Interviewing- The Essentials* edited by Michel St-Yves, published in 2014 by Carswell publications

Mark Handler

A topic intimately related to polygraph, and yet often overlooked, is Investigative Interviewing. Many polygraph examiners are unfamiliar with the concept. I hope by way of this book review to introduce interested readers to Investigative Interviewing. I can think of few better ways to familiarize oneself with the essentials of concept than by reading this book. The practice recommendations from this book will surely improve the quality of anyone's investigative work, regardless of the milieu.

The book is for anyone who interviews anyone else, but is especially appropriate for polygraph examiners who are often in the unique position of neutral fact-finder. Polygraph consumers and end-users look to polygraph to solve problems that for the moment seem unclear. There is no better time to take advantage of the essentials of Investigative Interviewing than during a polygraph pretest interview. Examiners are in a unique position to establish an information gathering environment. Examinees can be cajoled into providing information that can be exculpatory or inculpatory. The interview setting can give them rope to pull themselves out of their proverbial hole - or hang themselves in the process.

I often hear examiners say they "seek the truth", but that really isn't the first step in the process. We can't get to the slippery truth without facts to check out. We can't get the facts to check out if we don't interview. Information is the lifeblood of any investigation. It provides direction; it can show attempts at misdirection. Information helps confirm what we know, disconfirm what we thought we knew, and help reveal what we don't know. Having the tools to best develop information is essential- as the book points out.

Michel St-Yves is a Canadian forensic psychologist who works with the police. He is a friend and advocate for law enforcement which is clearly reflected in his work. His area of expertise is in conducting investigative interviews and teaching law enforcement to do so. He gathered some of the world's leading experts on the subject and had them write "how to" chapters for the book, geared towards the practitioner. While many books from academicians focus on theory, this is not one of them. This is for the investigative interviewer and it is especially relevant for polygraph examiners.

The book begins with a wonderful primer on rapport by St-Yves. Investigators are often taught the importance of establishing and maintaining rapport. But what does rapport look like? How do we get (and keep) it? Why is it essential to the investigative interview? The author tackles these, and many other of the thorny questions about rapport. In my opinion it would be worth buying the book for this chapter alone. For without rapport, the interview is doomed.

The second chapter is an update on the Cognitive Interview (CI), which should be used in every polygraph pretest interview. Ed Geiselman and Ronald Fisher developed the CI around 1985 and published their first book in 1992. In this chapter they describe updates, improvements and findings about the CI. The CI is a general strategy for guided memory retrieval based on scientific knowledge of human memory. The goal is to generate rich detail, without contamination. Over 100 empirical studies show a 25-50 % increase in detail over a standard police type question and answer interview.

In chapter three experts on child interviewing provide recommendations for conducting physical and sex abuse investigative interviews. Cyr, Dion and Powell break down

and discuss best practices for the child interview. They include; planning & preparation, communication rules for obtaining an account, establishing and maintaining rapport, memory limitation discussion and questioning strategies. The authors give several example interview protocols that have been scientifically shown to work well with children. Finally they remind us that children, including adolescents, are not just "little adults" and we need to modify our interviewing approaches to maximize information gain with this population.

Chapter four deals with eye witness memory and identification. Hope and Sauer are cognitive psychologists whose focus and expertise is human memory and decision making. We are asking examinees what they remember so it is incredibly important to have an understanding of the limits of human memory. Likewise, we engage them in decision making process, tell or don't tell, so we should have a basic understanding of neuroeconomics. Much of this chapter focuses on witness identification so it will be more useful to examiners who also conduct police investigations that include witness identification. The authors provide the current best practice standards for conducting show-ups, line-ups and photo identification. They also give sound advice on presenting witness identification evidence in court.

A short chapter on false memory by James Ost follows and reminds us that we have incredible power in the interview room that can create false memories. Ost is a false-memory expert who has published extensively on the subject. He provides a short review that includes; what are false memories, how do they occur, what are some of the mechanisms know to create them, what is the evidence of their existence? Most importantly he provides clues or indicators of concern that a reported memory may be false. Much of this relates to claims of physical abuse and sexual abuse reports, which constitute the bulk of many police polygraph examiner's workload.

Chapter six is the heart of the book, in my opinion. Written by Michel St-Yves and Christian Meissner, two of the current leading authorities on suspect confession and confession related concerns. They review; the importance of confession evidence, who confesses, why people confess, the internal and external pressures that precipitate confession, personality factors that affect confession, and much more. They break down "interrogation" into its component parts that mirror the P.E.A.C.E. model. They discuss important verification and control practices to try to ensure the confession is real, and not false. Many of us think we know what to do, what not to do, and how to do it. This chapter provides a benchmark against which to see if you are following recommended practices.

Gisli Gudjonsson is one of the world's leading authorities on mental vulnerabilities and false confessions. His chapter seven is a comprehensive review of the current state of knowledge on the subject. Mental vulnerabilities are psychological states and traits that increase a person's risk of providing inaccurate or unreliable information during an investigation. It includes; low intelligence, developmental disorders, personality disorders, high suggestibility or compliance, and recall concerns like memory distrust syndrome. Gudjonsson reminds us that just because someone has one of these, it does not invalidate their confession. His concerns and recommendations are prophylactic and protective. He reminds us how important it is to assess for vulnerabilities ahead of time, if possible. He also reminds us to reflect afterwards on whether the subject had any mental vulnerabilities that may have affected their statement or admission. Most polygraph examiners know that people we encounter can seem overrepresented by the groups most concerning to Gudjonsson. We can benefit by paying heed to his concerns and recommendations.

Aldert Vrij is a leading authority and researcher on the science of detecting deception. He has authored several books and numerous chapters and research articles on the subject. In chapter nine, Vrij updates us on "myths and opportunities in verbal and nonverbal lie detection". This chapter should be a must read for every police officer, police recruit, attorney, judge, and criminal justice professor or professional. Vrij summarizes the myths surrounding the unassisted human lie detector hypothesis. He provides examples of evidence-based practices that actually do separate truthful from deceptive subject, though

the differences are small. He closes out with a best practice recommendation for conducting an investigative interview.

The final chapter of the book is co-written by a number of experts in Investigative Interviewing training. They provide a framework for effective police interview training. They share their thoughts and experience on the best way to train new (and old) police investigators in Investigative Interviewing. Most of the authors have been involved in police training development for many years and have helpful insights on successful training strategies.

St-Yves skillfully closes the book with his concluding thoughts on the past and future of Investigative Interviewing. He reminds us that Investigative Interviewing has moved from the realm of art into the realm of science and art. Learning from scientists in allied disciplines will only improve what we do. There is a great deal of evidence from the lab and field that supports these recommendations. St-Yves recaps the essentials of good communication skills and their importance when interviewing witness, complainants and suspects. He reminds us that all good interviews require preparation and a mindset towards unbiased information gathering.

As polygraph examiners it seems we should be ethically bound to conduct investigative interviews during our pretest interview. We have a unique opportunity to gather information before conducting any test. That information can be exculpatory, inculpatory, or inform the investigation in some important way. If we approach the pretest interview as an information gathering event we can increase the information gained during the testing process. Many interviewees (truthful and deceptive) will provide information during an appropriate investigative interview. They simply have to be given the opportunity. The book *Investigative Interviewing - The Essentials* can open your eyes to a world of improvement. I seldom recommend any book with such enthusiasm, but this is a rare occasion.

The French novelist Marcel Proust said, "The real voyage of discovery is not in seeking new landscapes but in having new eyes." I hope this book helps you see interviewing through new eyes. It certainly did so for me.

Police Cadet Attrition and Training Performance Outcomes

Adam Park

Texas Department of Public Safety

James S. Herndon

Private Practice, Orlando, Florida

Abstract

Research devoted to the predictive validity of criteria commonly used to screen police applicants has received little attention. The need has increased for police agencies to evaluate their various screening methodologies in the multiple-hurdles approach to police candidate selection. Grounded in Schmidt and Hunter's theory of general mental ability in job performance, this study examined the predictive validity of candidates' demographic profiles and results of the pre-academy screening polygraph to predict training outcomes and attrition. This quantitative study used logistical and linear regression analysis to determine whether these variables were viable screening mechanisms to predict attrition and training performance among police cadets at the Texas Department of Public Safety. Each independent variable (age, prior military service, level of education, and polygraph result) predicted cadet academy completion status (unsuccessful; successful). However, there was no evidence to suggest that age, prior military service, or level of education predicted training performance as measured by final academy grade point average or score on the Texas Commission on Law Enforcement Officer Standards and Education (TCLEOSE) exam. This study's findings relative to each independent variable support contemporary police research by identifying potentially valued characteristics of a successful police candidate. Additionally, these findings could allow police administrators to better implement training strategies that compliment agency goals; thus better preparing candidates to protect society. By understanding the validity of these screening procedures in candidate selection, police agencies could save time and money.

Keywords: *cadet, attrition, polygraph, screening, demographic profile*

Police Cadet Attrition and Training Performance Outcomes

From a practical point of view, the most valuable determinant in training or job performance is the predictive validity of essential work functions (Schmidt & Hunter, 1998). However, in an ever-changing world, it is becoming increasingly difficult to find quality personnel for the police profession (Henson, Reyns, Klahm, & Frank, 2010). Additionally, scholars continue to debate which variables are the best predictors of quality personnel (Henson, et al., 2010). This leaves a gap with respect to which other variables consistently predict the best cadet. Given the existing array of selection criteria police agencies use in the preemployment process (e.g., age, prior military service, psychological and physical agility testing), it is important to examine their relationship to police cadet performance.

For more than 90 years researchers have studied psychological test data in the

Author Note

The views expressed in this paper do not necessarily reflect those of the Texas Department of Public Safety. Correspondence concerning this article should be addressed to Adam Park Ph.D., Cypress, TX 77429.

Contact: hyperion1975@aol.com

context of selecting law enforcement personnel (e.g., Terman & Otis, 1917); yet there exists some disagreement regarding the issue of best practices in the "multiple hurdles" approach to police cadet selection. Today, most candidates for the position of police cadet must, as a part of the screening process, participate in a background investigation and submit to psychological and polygraph tests (Fuss & Snowden, 2004). Contemporary screening methods also generally include written tests, physical agility tests, and meeting certain demographic thresholds (Decicco, 2000). These methods have dual functions: to select the most competent candidate for training as well as vet the candidate to ensure that agency goals are met and public safety is not compromised (Decicco, 2000).

Many organizations employ various selection techniques in the preemployment process, and often take action relative to the weights given to each point of view (Society for Industrial and Organizational Psychology, 2003). What this means for the cadet selection process is that the police agency is often burdened with the cost in time and money of using numerous techniques. Questions remain about which predictors correlate best with police cadet performance (White, 2010). The majority of psychological screening for police cadets in the United States uses personality measures (Hancock & McClung, 1987). Recent studies argue for the use of non-clinical personality assessments for police cadet selection (Forero, Pujol, Olivares, & Pueyo, 2009); while other studies suggest that demographics and residency have little to do with overall cadet success (White, 2010).

Important to this discussion is a strategy that effectively ties selection mechanisms that are part of the demographic and pre-academy screening data to cadet attrition and training performance. This, it was suggested, would provide a better bridge to existing disparities in police candidate selection research. There is a gap in the empirical research that connects the predictive validity of each predictor in this study to cadet attrition and training performance. Additionally, little research has been devoted to using multiple predictors to analyze police academy performance; with current efforts leaning heavily toward using psychological traits to predict the job performance of a police officer (White, 2008). Noting the importance of finding plausible links to the multiple predictors selected in this study, it is appropriate to make a brief introduction to the multiple hurdles approach to general and police employee selection (e.g., application, demographic characteristics used in the hiring process, interview, background investigation, psychological testing, and integrity testing). Additionally, a brief discussion of the history of the Texas Department of Public Safety (DPS) will provide the context for using the demographic data in this study.

Literature Review

The Texas DPS uses a non-compensatory model-better known as a successive or multiple hurdles approach to police cadet selection, where the applicant must pass a series of tests (Henson et al., 2010). For example, failing the physical fitness test eliminates a candidate from further consideration; the same is true of failing or performing below established thresholds in any screening process (medical, psychological, polygraph). In its equation for selecting cadets, the Texas DPS uses certain demographic criteria (minimum age, prior military service, education). There is no ceiling on age, but applicants must be at least 20 years of age upon graduation from the academy (Texas DPS, 2012). Applicants bring a mixture of experiences to the preemployment process. The Texas DPS addresses the issue in a number of ways. While applicants for the position of police cadet at the Texas DPS must have earned at least 60 credit hours from a regionally accredited college or university (Texas DPS, 2012), they can substitute prior military or law enforcement service for college education, equivalent to the 60-hour requirement.

Police agencies use many screening items in both the preemployment and training academy process (White, 2010). Harris, Dworkin, and Park (1990) examined the predictive validity of numerous screening procedures in the hiring process for large public companies. The three notable screening procedures reported as most accurate in selecting future superior workers were targeted interviews, accomplishment tests, and references. In contrast, physical ability tests, polygraph tests, and genetic tests were deemed most inaccurate in the selection of a superior worker; however, failed drug and polygraph tests were two of the most likely reasons employers gave that would affect hiring decisions (Harris et al, 1990). In order to provide a more broad understanding of applicant selection processes, a portion of the screening procedures utilized in previous studies incorporating both private companies and police agencies are included as predictors in the present study. For example, acade-

my tests, as well as the state licensing exam were utilized as predictors for accomplishment tests. Another study noted the work sample is considered to have a high degree of validity because of its practical nature (Osoian, Zaharie, & Lazar, 2011), while other employers look to cognitive ability tests as a predictor of hiring success (Lewis, Shimerda, & Graham, 1983).

Age as a Predictor of Training Performance

Perhaps no predictor in this study has been underrepresented in empirical police cadet selection research than that of age. There have been numerous attempts to explain age-related cognitive changes such as common cause theory, the processing speed theory, and the executive function hypothesis (Luszcz & Byran, 1999); however, these efforts have been difficult due to the various differences in statistical findings which make generalizations about age across time arduous at best (Clay et al., 2009). The specific focus in this study was to examine age as a predictor of cadet attrition and training performance using the theory of General Mental Ability GMA. Therefore, in this study, GMA would be represented by the cadet's c academic test results in a 28- week police academy. Age has long been judged a great variable in organizational contexts due to its common relationship to the goals and objectives of many agencies (Randhawa, 2007). Obvious correlations exist utilizing age as a predictor of physical performance in any setting; however, this study does not examine physical fitness as a predictor of training performance.

It is appropriate to now examine previous studies to assist in establishing a background perspective of age's relationship within law enforcement training contexts. Chappell (2008) used age, race/ethnicity, gender, military experience, education, special position (ranking officer in class), and type of academy as independent variables to assess cadet performance in a re-designed academy (community policing) versus a traditional academy setting. A literature review revealed one study that included age and training outcomes. White (2008) examined a large metropolitan police training class (n = 1,556) and found that as a cadet's age increases, their academy performance decreases. The practical value of age as it relates to police training appears to be largely unknown, thus making it necessary to examine this predictor's current status in police screening contexts.

Biographical data has been substantially correlated with GMA (Schmidt, 1988). Some scholars offer that police agencies prefer younger applicants because they are less rigid and more accepting of the community-based policing model recently adopted by many police agencies (LaRose, Caldero, & Gutierezz, 2006). Aamodt's (2004) review of 300 theses, dissertations, journal articles, and conference papers yielded several studies addressing academy performance, but included age only as a sample characteristic. However, recent research covering broad spectrum occupations (i.e. occupations other than public safety) asserts that workforce age is linked negatively to quantitative organizational performance, but positively associated with qualitative performance (Gellner, Schneider, & Veen, 2011). In other words, output appears to be correlated with being younger, whereas an older worker shows to produce more quality in their product. One viewpoint on age as it relates to police performance posits a likely paradigm shift from hiring younger police applicants, to recruiting older prospects. For example, some police departments have sought out second-career applicants from other professions not because of their previous law enforcement experience, but due to their maturity and documented stability (Bennett & Hess, 2004). Other studies examining age and work performance report that demographics point to an older workforce (Sharit, Czaja, Hernandez & Yang, Y, 2004). As it has been suggested, the reports in this area are dubious and provide for a perplexing understanding of age and police training performance.

The Utility of Prior Military Service

Police training programs have tended to follow militaristic patterns, which explains, in part, rationales for police agencies hiring former military members as patrol officers (Birzer, 2003). Historically, there has been a strong representation of individuals in law enforcement with prior military experience (Aamodt, 2004). Evidence of this can be seen in the roots of many police units throughout the United States. For example, police agencies adopted quasi-military models in the early 1900s in efforts to eliminate corruption (Fogelson, 1977).

Despite strong academic opinions, scholars continue the debate regarding the practical value and contextual meaning of the word "paramilitary". Law enforcement agencies traditionally have adopted a "paramil-

itary" model, despite a lack of empirical evidence supporting the utility of a military structure within law enforcement contexts (Bittner, 1990; Franz & Jones, 1987). Some scholars debunk the notion of police as a paramilitary organization (Cowper, 2000), while others agree the two share commonalities, but debate about a common definition of paramilitary when applied to a law enforcement environment (Jefferson, 1987, 1993; Waddington, 1993). Recent findings on the nature of policing suggest there appears to be a convergence of roles between police and military function (Campbell & Campbell, 2010). Prior research reveals positive correlations with police work performance, but there remains a void with respect to the usefulness of military service to police training.

College as a Determinant of Success

It is becoming more common for applicants who enter the police cadet selection process to have some level of college education, as studies reveal that 44% of police applicants have attended at least 1 year of college (Bennett & Hess, 2004). In contrast to the previous discussion on military service, statistics reveal that police applicants now have more college education than military background (Bennett & Hess). Using college experience as a predictor variable is an important component of this study given its unknown predictive validity in police training contexts. Another reason for the examination of college education as a predictor lies with the fact that the literature has produced mixed reviews with respect to correlating formal education and training success (Walker, 1994).

Criminal justice education has typically focused on a distinct, three-level system (high school diploma, associate's degree, and bachelor's degree) that in many ways ties into entry-level law enforcement jobs (Buerger, 2004). As a practical matter, the previous statement assists in establishing criteria for incorporating the three-level system (high school diploma, associate's degree, and bachelor's degree), with the addition of some college experience, as predictors. The practical value of having a college degree in the police profession is still largely misunderstood (Walker, 1994). One rationale for this uncertainly is grounded in the thought that individuals make decisions relative to pursing higher education on the foundation of a cost-benefit analysis (Brand & Xie, 2010).

As early as 1916, August Vollmer, the father of modern policing, underscored the importance of education for officers (Guthrie, 2000). The Wickersham Commission (1937) and the President's Commission on Law Enforcement and the Administration of Justice (1967) highlighted the significance of a post-secondary education for police officers (Bennett & Hess, 2004).

Although the previous discussion noted that more applicants come to the screening process with some level of college education, studies show that this is certainly not an expectation (Capsambelis, 2004). A study from the Bureau of Justice Statistics noted that nationally only 1% of police agencies required a 4-year degree; whereas only 6% required some college, and 8% mandated a 2-year degree (Paoline & Terrill, 2007). Advocates of the college experience argue that educated officers produce better reports, receive fewer complaints, and produce a better overall work product (Baker, 1995; Carter et al., 1989; Trautman, 1986; Vodicka, 1994). Other scholars stated that the college experience provides a better opportunity for an individual to mature, offers a broader base of general knowledge, and enhances verbal and communication skills (Armstrong & Polk, 2002). Taking this evidence into account, administrators would perhaps suggest the college experience a valuable criterion for cadet selection.

The TCLEOSE Exam

In 1965 the Texas State Legislature created the Texas Commission on Officer Standards and Education (TCLEOSE) to establish standards for peace officers (TCLEOSE, 1997). The Basic Peace Officer Course Format consists of 618 hours of academics related to entry-level policing and is a test of job content knowledge (TCLEOSE, 2008). The TCLEOSE Exam is comprised of 250 multiple choice questions that address: (a) Texas Penal Code; (b) Texas Code of Criminal Procedure; (c) The Texas Constitution; (d) Texas Traffic Law; (e) Drug Questions; (f) Police approaches to family violence and mental health; and (g) Civil Law (TCLEOSE, 2010). Applicants for the position of police cadet, who are not already peace officers, must first pass the TCLEOSE Exam before becoming commissioned troopers (Texas DPS, 2012). This study incorporated only datasets from subjects who had not previously taken the TCLEOSE exam.

A comprehensive literature review revealed that great disparity exists in the realm of research related to the validity of state police licensing examinations. Although not a civil service test, the TCLEOSE exam serves to

assess the cadet's knowledge of general police aptitude. Schroeder (1973) studied the validity of the entrance examination for the position of patrolman under the guidelines established by the Equal Employment Opportunity Commission (EOEC) and found exam scores were positively related to performance. A review of literature relating to the validity of the TCLEOSE exam yielded no results. However, it was discovered Peace Officer Standards and Training (POST) commissions are located in every state so as to set minimum requirements for entry-level law enforcement positions (Bennett & Hess, 2000).

The Polygraph

Although empirical studies fail to investigate polygraph's predictive validity for-training performance in the screening process for officers, literature suggested the method of preemployment polygraph is only increasing (Krapohl, 2002). Additionally, there is strong evidence that integrity tests have practical application when paired against cognitive ability (Schmidt & Hunter, 1998). For example, GMA has produced more incremental validity regarding the prediction of private employee training performance than any other measured study to date (Schmidt & Hunter, 1998). Meesig and Horvath (1995) reported that 99% of large and 90% of small law enforcement agencies require the use of a polygraph as a condition of employment for sworn positions. The reliability of a candidate's truthfulness is of high value to police administrators throughout the police preemployment process. A meta-analysis of integrity test validities found that preemployment tests of honesty can predict certain organizational disruptive behaviors (Ones, Viswesvaran, & Schmidt, 1993). Extending on this idea, the prevalence of general preemployment polygraph screening appears to be on the rise. The use of the polygraph is prohibited in most private sector arenas because of the Employee Polygraph Protection Act of 1988 (EPPA) (Decicco, 2000). Platform for the inclusion of polygraph results in this study lies in the fact that virtually nothing is known about the existing relationship between polygraph results in preemployment settings and training performance in police cadets or officers. This is the first known study to incorporate polygraph examinations as a predictor of police cadet attrition and cadet performance.

One study (Ho, 2001) was found to have utilized polygraph results as one predictor for assessing the effect of a psychologist's recommendations for hiring. Ho (2001) used linear regression to examine the effects of independent predictors (demographics, gender, age, prior military service, self-reported drug usage, and prior encounters law enforcement) on each dependent measure. Past research has even examined the correlation between civilian preemployment tests and future employee behavior. Although much controversy still exists surrounding the validity of polygraph, many police departments today use this instrument as a tool for veracity in the officer selection process (e.g., Ben-Shakhar & Furedy, 1990; Lykken, 1981; Saxe, 1994). This would imply, in a purely non-systematic way that police administrators have found a certain value to the requirement of a polygraph as a condition of employment. More research is needed in the area of preemployment polygraph to facilitate knowledge with respect to its utility in these settings.

Theoretical Base

The theoretical base for this study was grounded in the framework of predicting occupational performance and the validity of paired combinations of general mental ability (GMA); defined as the outcome of GPA and scores from the TCLEOSE exam. This theory was introduced in 1904 by C. Spearman and, as Schmidt and Hunter (2004) noted, is often used for predicting occupational performance: "GMA predicts both occupational level attained and performance within one's chosen occupation and does so better than any other ability, trait, or disposition" (p. 162). According to Schmidt and Hunter (1998) the percentage of validity for preemployment personnel measures (i.e., integrity tests, biographical data measures, and years of education) increase together. The thrust of this study was to examine this theory by assessing GMA's ability to predict training performance.

There are two reasons for using GMA: (a) it has the highest level of validity in personnel selection and the lowest cost in terms of monetary measurement (Schmidt & Hunter, 1998) and; (b) it repeatedly provides the best evidence of validity among other measures (Hunter, 1986; Hunter & Schmidt, 1996; Ree & Earles, 1992; Schmidt & Hunter, 1981). This evidence makes GMA a viable avenue for future research on selecting cadets. In the present study, grade point average (GPA) and TCLEOSE exam scores were reported as a measure of GMA (e.g. cadet GPA at end of academy and TCLEOSE score). From a theoretical perspective, the utilization of GMA,

along with multiple predictors (i.e. age, prior military service, and education) was the best choice, considering the similarity of items used in Schmidt and Hunter's (1998) work and the screening methods used by the Texas DPS in the preemployment process.

The theory of GMA embraces general intelligence and specific aptitudes and abilities; it then shows important differences between groups (Schmidt & Hunter, 1998). This was important considering the proposed research questions in this study. Finally, elements that comprise GMA have been examined in both military and police occupations, thus supporting the idea of using the theory to predict training performance in a police academy. Grade Point Average (GPA) and the Texas Commission on Law Enforcement Officer Standards and Education (TCLEOSE) exam score were considered outcome or criterion variables since they were generated by the applicant throughout the police academy training.

The purpose of this quantitative study was to examine whether the two sets of variables, demographic profiles and pre-academy polygraph screening results, were significant predictors of police cadet attrition and training performance at the Texas DPS. It incorporated data from both the preemployment process and the training academy. Many studies have generalized various independent and dependent variables in attempts to correlate demographic data with training and job performance (Aamodt, 2004). Some studies have used biographical data with test data, while others have mixed interview data with archived data. However, no study was found in the literature that examined the predictive validity of age, education, academic performance, and polygraph results to predict cadet attrition and training performance.

Demographic variables are aspects of the applicant they bring into the selection process. For example, demographic variables in this study were age, prior military service, and level of education. A fourth predictor, preemployment polygraph results, was analyzed to assess the polygraph's ability to predict attrition as well as training performance. The polygraph result was incorporated as a predictor to determine its usefulness for predicting attrition rates among police candidates.

All candidates were administered a polygraph examination before admission into the academy. As per DPS policy, candidates determined to be either inconclusive or deceptive were interviewed and administered a second "break-out" examination. This process is established to assist the candidate in clearing inconsequential issues related to hiring, or to obtain additional information that may disqualify the candidate from the hiring process. Candidates receiving a second deceptive or inconclusive result were disqualified from the process. Those candidates who were administered a second break-out polygraph and scoring no deception were admitted to the academy. Because a portion of the candidates in the first inconclusive and deceptive group were ultimately cleared and admitted to the academy, they were separated from the first no deception first polygraph test group. This was considered by the researcher to be of value in determining the group's predictability in training success as opposed to the group who received no deception on the first examination.

It was hypothesized that younger cadets who possessed a college degree and prior military experience would perform better in a training academy environment as measured by GPA and the TCLEOSE exam score, as opposed to older cadets who possessed no formal college education or prior military service. Finally, it was hypothesized that cadets entering the academy with a polygraph result of "no deception indicated" on their first polygraph examination had a better chance of completing the academy, as opposed to those cadets having either a first examination result of "inconclusive" or "deception indicated"; thus requiring a subsequent "break-out" examination.

Hypotheses:

H_1^1: The demographics of cadet age, military experience, and level of education are predictive of academy completion for cadets who participated in the 2008 Texas Department of Public Safety Training Academy.

H_1^2: The demographics of cadet age, military experience, and level of education are predictive of overall training performance for cadets who participated in the 2008 Texas Department of Public Safety Training Academy.

H_1^3: Preemployment polygraph results of "No Deception Indicated" are predictive of academy completion for cadets who participated in the 2008 Texas Department of Public Safety Training Academy.

Methodology

This quantitative study used a non-experimental, descriptive design. Cadet demographics and pre-academy polygraph screening results were captured from Texas DPS records. Categorizing these predictor variables was consistent with Aamodt's (2004) research which examined each variable used in this study, other than polygraph test results, to police officer training and job performance. The demographics of the population under examination provided for a largely unified sample. This study included stratification of the population; incorporating true characteristics (e.g., White/Hispanic males) representative of the entire sample (Fowler, 2002). Specifically, 195 subjects were examined to find if demographic profiles and pre-academy polygraph screening results were significant predictors of attrition and training performance by the selected population. Eligibility for group classification included: (a) Each participant was to be an applicant for the position of police cadet at the Texas DPS for the 2008 academy; (b) The participant must have completed demographic information which documented age, any prior military service, and level of education as criteria for selection; (c) The participant had to numerically score no deception indicated, inconclusive, or deception indicated on a first or second attempt of the Texas Department of Public Safety Modified General Question Technique (DPSMGQT); and (d) The participant must have either failed to complete training after beginning the academy, or graduated from the Texas DPS academy with numerical scores for grade point average (GPA), as well as participated in a first attempt of the TCLEOSE exam, which provided a numerical score. Cadets are allowed to take the examination multiple times; therefore, the researcher included only numerical scores from the first attempt of the TCLEOSE exam so as to better compliment true test-taking ability. Prior police officers who held a TCLEOSE license at the time of training were not included in any dataset; as they were exempt from the state examination.

Instrumentation

Demographic data was derived from the preemployment polygraph questionnaire form (HR-39). Outcome variables were collected by obtaining scores from each cadet's cumulative grade point average (GPA) and score on the Texas Commission on Law Enforcement Officer Standards and Education (TCLEOSE) exam. Finally, pre-academy screening polygraph results were gathered from data produced on the Texas DPS Modified General Question Technique (DPSMGQT). As previously discussed, the TCLEOSE exam is comprised of 250 multiple choice questions and covers the topics of: (a) Texas Penal Code; (b) Texas Code of Criminal Procedure; (c) Texas Constitution; (d) Texas Traffic Law; (e) Drug Questions; (f) Police approaches to family violence and mental health and; (g) Civil Law (TCLEOSE, 2012). The polygraph instrument used for the collection of all physiological data was the Axciton Five-Channel Computerized Polygraph System. Channels within polygraph contexts refer to the individual components attached to the examinee. The data gathered from polygraph examinations administered in this study was exclusively archived. Those examinations were conducted at two locations: (a) The Texas DPS Headquarters in Austin, Texas- Physical Address 5805 North Lamar Boulevard, and (b) Building M at the Headquarters Complex in Austin, Texas.

Analysis

This study was driven by past methodology, as a number of studies have incorporated regression analysis in attempts to assess cadet training outcomes (Aamodt, 2004; Guller, 2003; Jacobs & Solomon, 1977; Sanders, 2008; Waugh, 1996), though no known study existed that incorporated the multiple screening criteria in this proposal in attempts to predict cadet attrition and training performance. Research consistently shows that regression analysis is the most appropriate statistical treatment with respect to predicting police academy performance, as it permits the relationship between variables to be inspected due to the correlation of variables (Bernstein, Schoenfeld, & Costello, 1982; Waugh, 1996). The core of the analysis was linear and logistical regression for all three hypotheses. Regression analyses are used to predict the relationship between predictor variables and outcome or criterion variables (Tabachnick & Fidell, 2001). More specifically, regression analyses are used to study the relationship of a dependent variable y to two or more independent variables, using a regression model that is represented by the equation $y = ß0 + ß1x1 + ß2x2 + \varepsilon$. This model also includes ε, or an error variable, which is a random variable that refers to the variability in y that the listed independent variables do not account for.

Training performance (GPA and TCLEOSE exam score) and cadet attrition were treated as two separate analyses. Addition-

ally, GPA and TCLEOSE results were not reported for a cadet that failed to complete the academy. The rationale behind this methodology was based on the premise that a cadet would be required to graduate in order to capture a cumulative GPA and qualify to sit for the state licensing exam. Within this context, a cadet selected to the academy that ultimately resigned for undisclosed reasons or failed out was included in the polygraph dataset, as this supported the hypothesis for cadet attrition. Cadets that were expelled or resigned were treated as analogous.

Results

Among the 195 study participants, 190 (97.4%) were male and 5 (2.6%) were female. The ethnic distribution was 132 (67.7%) White; 56 (28.7%) Hispanic; and 7 (3.6%) Black. The education distribution was 48 (24.6%) high school graduate or possessing a GED; 44 (22.6%) with some college; 40 (20.5%) possessing an associate's degree, and; 63 (32.3%) holding a bachelor's degree. The average (and standard deviation) age was 27.8 (6.4) and the range was 20 to 52. There were 121 (62.1%) study participants that had no prior military service and 74 (37.9%) with prior military service. A total of 150 (76.9%) study participants completed the academy and 45 (23.1%) failed to complete the academy. The average (and standard deviation) academy GPA was 92.1 (3.35) and the range was 83.1 to 97.6. The average (and standard deviation) TCLEOSE exam score was 80.2 (6.2) and the range was 67 to 93. Polygraph decisions were used as the measure for inter-rater reliability. Cohen's Kappa statistic *K* to measure inter-rater reliability of the polygraph test (n = 10 polygraphs; 2 raters) was .40 (moderate agreement). A total of 121 (62.1%) study participants had a polygraph result of "No deception indicated" and 74 (37.9%) "Inconclusive or deception indicated" (see Tables 1 - 3).

Table 1 Descriptive Statistics for Ethnic Distribution

	Frequency	%	Valid percent	Cumulative %
White	132	67.7	67.7	67.7
Hispanic	56	28.7	28.7	96.4
Black	7	3.6	3.6	100.0
Male	190	97.4	97.4	
Female	5	2.6	2.6	
Total	195	100.0	100.0	

Note: Among the 195 study participants, 190 (97.4%) were male and 5 (2.6%) were female. The ethnic distribution was 132 (67.7%) Caucasian; 56 (28.7%) Hispanic; and 7 (3.6%) Black.

Table 2 Descriptive Statistics for Age by Academy Completion Status

Academy Completion Status	N		Mean	SD	Minimum	Maximum
	Valid	Missing				
Unsuccessful	45	0	29.67	6.779	21	52
Successful	150	0	27.19	6.147	20	52

Note: The average age was significantly smaller for those who completed the academy compared to those who failed the academy. The average (and standard deviation) age was 29.7 (6.8) versus 27.2 (6.1) for those who failed the academy and those who completed the academy, respectively, $t(193) = 2.31$; $p = .022$. This finding was consistent with the findings from hypotheses 1 and 3 where older age was associated with lower odds of completing the academy.

Table 3 Frequency Distribution of Study Participant's Polygraph Result

	Frequency	Percent
No Deception Indicated	121	62.1
Inconclusive or Deception Indicated	74	37.9
Total	195	100.0

Note. Ten polygraph examinations were scored by two examiners. Their scores were then utilized to calculate inter rater reliability using the *K* statistic.

Hypothesis 1 was tested using stepwise multiple logistic regression analysis. The dependent variable was academy completion status (successful; unsuccessful). As Table 4 illustrates, all three independent variables were found to be statistically significant. This means that prior military service ($p = .004$), age ($p = .006$), and level of education ($p = .011$) provided independent information in predicting academy completion status. That is, the three independent variables explained independent variance in academy completion status. The Nagelkerke R-Square Statistic associated with prior military experience was .05; .054 for age, and; .045 for level of education.

Table 4 Stepwise Multiple Logistic Regression Analysis of Academy Completion versus Age, Prior Military Service, and Level of Education

		b	SE	Wald	df	p-value	OR [e]	95% C.I. for OR	
								Lower	Upper
Model [a]	AGE [b]	-.078	.028	7.617	1	.006	.925	.875	.978
	PMS [c]	1.232	.423	8.485	1	.004	3.429	1.497	7.857
	EDU1 [d]	-1.017	.400	6.448	1	.011	.362	.165	.793
	Constant	3.289	.829	15.729	1	.000	26.816		

Note. The most important predictor of completion status was prior military service, age, and finally level of education. The three independent variables collectively explained 10.4% of the variance in completion status.

Hypothesis 2 was tested using stepwise multiple linear regression analysis. There were two separate measures of training performance (dependent variables): (1) Academy GPA, and; (2) TCLEOSE exam score. Therefore, this analysis was repeated for each of the two measures. For the first regression analysis, the dependent variable was the academy GPA. The independent variables entered into the stepwise model selection procedure were age (measured on a continuous measurement scale in years), prior military experience (0 = No; 1 = Yes), and; level of education. As was done in testing hypothesis 1, education was re-coded into dummy variables prior to conducting the analysis. None of the independent variables met criteria for entry into the model (i.e. $p < .05$). As Table 5 illustrates, the null hypothesis was not rejected and it was concluded that neither age, prior military service, or education level were predictive of academy GPA. The *p*-values for each of the independent variables were .96 for age; .68 for prior military service; .86 for EDU1 (high school diploma); .96 for EDU2 (some college) and; .19 for EDU3 (college degree).

Table 5 Stepwise Multiple Linear Regression Analysis of Academy Grade Point Average [a] versus Age, Prior Military Service, and Level of Education

	Unstandardized Coefficients		Standardized Coefficients		
Model [a]	b	SE	Beta	t	p-value
(Constant)	91.729	1.351		67.884	.000
AGE [b]	.002	.047	.004	.048	.962
PMS [c]	-.261	.639	-.039	-.409	.683
EDU1 [d]	.156	.860	.018	.181	.856
EDU2 [e]	.043	.861	.005	.050	.960
EDU3 [f]	1.062	.805	.151	1.320	.189

Note. It was found that neither age, prior military service, or education level were predictive of academy GPA. Age = cadet age (measured on a continuous measurement scale in years); PMS =prior military service; EDU = level of education; b = Estimated values of raw (unstandardized) regression coefficients; SE = standard error.

For the second regression analysis, the dependent variable was the academy TCLEOSE exam score. The independent variables entered into the stepwise model selection procedure were age (measured on a continuous measurement scale in years), prior military experience (0 = No; 1 = Yes), and; level of education. As was done previously, level of education was re-coded into dummy variables prior to conducting the analysis. Table 6 shows that only EDU2 (some college experience) met criteria for entry into the model, $p = .023$. The null hypothesis was not rejected and it was concluded that age, prior military service, and level of education do not explain independent variance in TCLEOSE exam scores. The equation of the model is: TCLEOSE = 80.82 -2.82*EDU2, where TCLEOSE = the average Texas Commission on Law Enforcement Officer Standards and Level of Education Score; EDU2 = level of education (0 = Not Associate's degree; 1 = Associate's degree).

Table 6 Stepwise Multiple Linear Regression Analysis of TCLEOSE [a] Exam Scores versus Age, Prior Military Service, and Level of Education

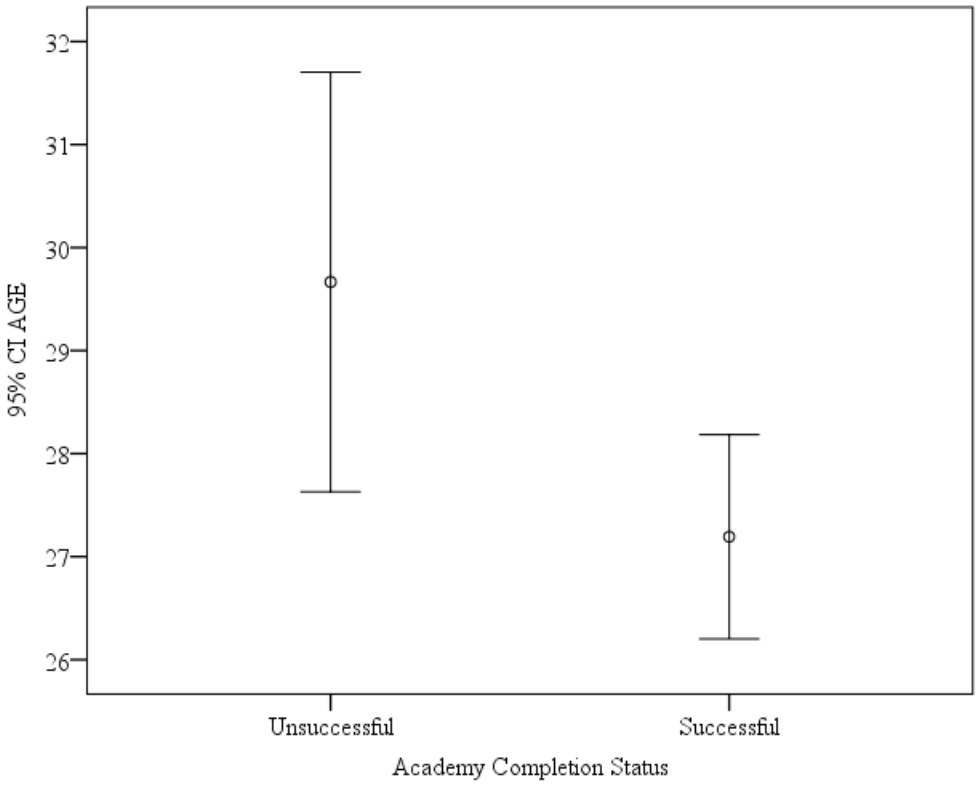
Model [a]	Unstandardized Coefficients		Standardized Coefficients		
	b	SE	Beta	t	p-value
(Constant)	80.822	.566		142.788	<.001
EDU2 [b]	-2.822	1.225	-.186	-2.303	.023

Note. Only EDU2 (some college experience) met criteria for entry into the model, $p = .023$. The null hypothesis was not rejected and it was concluded that age, prior military service, and level of education do not explain independent variance in TCLEOSE exam scores. *b* = Estimated values of raw (unstandardized) regression coefficients; *SE* = standard error; *t* = sample value of the *t*-test statistic; *p*-value = probability value.

Hypothesis 3 was tested using simple logistic regression analysis. The dependent variable was academy completion status (successful; unsuccessful). The independent variable was the initial polygraph test result ("No deception indicated"; Inconclusive or Deception Indicated"). The null hypothesis was rejected and it was concluded that polygraph test results are predictive of academy completion (see Table 4). The Nagelkerke *R* Square statistic was .044, which means that the polygraph test results explained only 4.4% of the total variance in academy completion status.

Considering that the results for testing Hypothesis 1 showed that age, prior military service and level of education were predictive of academy completion status, it was of interest to determine if the polygraph test result explained additional variation in academy completion status, above and beyond the variance explained by the three demographic variables. A stepwise multiple logistic regression analysis was performed in order to address this. The independent variables entered into the stepwise model selection procedure were age, prior military experience, initial polygraph test result, and level of education. All four independent variables were statistically significant. This means that polygraph test result ($p = .011$), age ($p = .003$), prior military service ($p = .003$), and education level ($p = .011$) provided independent information in predicting academy completion status. That is, the four independent variables explained independent variance in academy completion status. The Nagelkerke *R* Square statistic was .054 for age; .050 for prior military service; and .048 for polygraph test result, and; .043 for level of education. Thus, the most important predictor of completion status was age, followed by prior military service, polygraph test result, and level of education. The four independent variables collectively explained 19.5% of the variance in completion status.

Figure 1 illustrates the age disparity between successful and unsuccessful cadets.



Exploratory Analysis

The results of testing Hypotheses 1 and 3 showed that age, prior military service, level of education, and polygraph test results contributed independent information in predicting academy completion status. In order to further explore the relationships between the independent and dependent variables, bivariate analyses were conducted. A two-sample *t*-test was used to compare the average age between those who did, and did not, complete the academy. Chi-square tests were used to evaluate the relationships between prior military service, education level, and polygraph test results, and the dependent variable, academy completion status.

The average age was significantly smaller for those who completed the academy compared to those who failed the academy. Figure 1 is an error bar chart that shows the average and 95% confidence interval for the average age by academy completion status. The figure gives strong evidence that those who completed the academy tend to be younger on average, compared to those who fail the academy. The profile for Figure 1 illustrates the 95% confidence interval for age, as well as distributes the two categories of cadet age by completion status. Figure 1 clearly reveals disparity between the two status groups. The average (and standard deviation) age was 29.7 (6.8) versus 27.2 (6.1) for those who failed the academy and those who completed the academy, respectively, $t(193) = 2.31$; $p = .022$. This finding was consistent with the findings from hypotheses 1 and 3 where older age was associated with lower odds of completing the academy.

A chi-square test was performed in order to determine if there was an associationbetween level of education and academy completion status. There was not a statistically significant difference in the percentage of cadets that completed the academy among the four education groups, X^2 (3) = 5.85; p = .12. Although not statistically significant, the largest standardized residual (in absolute value) was 1.8, which shows the "some college" group contributed the most to the magnitude of the Chi-Square statistic. In particular, the percentage of cadets with "some college" that completed the academy (63.6%) was less than the percentage that completed the academy among the other 3 education groups (79.2% to 82.5%). This finding was consistent with the results from hypotheses 1 and 3 where the "some college" group was found to have lower odds of completing the academy compared to the other education levels. Specifically, age, prior military status and polygraph test results explained some of the variance in academy completion status. The residual variance (the variance left over) could be better attributed to other factors (i.e. level of education).

A chi-square test was performed in order to determine if there was an associationbetween polygraph test results and academy completion status. There was a statistically significantly smaller percentage of cadets that completed the academy with initial "Inconclusive or deception indicated" polygraph results, compared to those with a polygraph test result of "no deception indicated". The number (and percentage) of cadets that completed the academy was 100 (82.6%) versus 50 (67.6%) for the "No deception indicated" and "Inconclusive or deception indicated" groups, respectively, X^2 (1) = 5.88; p = .015. This finding was consistent with the findings from Hypotheses 1 and 3 where a polygraph test result of "inconclusive or deception indicated" was associated with lower odds of completing the academy.

Conclusion

There is an increasing body of research within police personnel selection; however, gaps remain as to what predictors tend to make the best cadet. Extending on the aforementioned statement, future research should consider how the process of training translates into making a better police officer. Physical fitness should be examined from the perspective of its impact to cadet attrition. Within this context, age would, on the surface, appear to hold some validity when considering high rates of cadet fallout. Anecdotal evidence suggests younger cadets would have less difficulty than older cadets in completing the physical challenges associated with state police academies. Research should consider peripheral analyses that incorporate academics and level of education at the time of attrition. Including this analysis might provide independent information as to which characteristics enhance cadet successfulness, as well as provide answers to the role of general intelligence in cadet attrition. Military experience has been shown to better prepare individuals for police academy settings (Campbell & Campbell, 2010). However, studies must be conducted that examine the military occupational specialty's (MOS) role in training (i.e., identifying which MOS better prepares an individual for a police academy). Additionally, it would be of interest to know which military

service (i.e. Air Force, Army, Marine Corps, or Navy) produces the best police cadet. Isolating these variables could assist military personnel in their transition out of the service and into the United States workforce. Police administrators could find such data useful in future recruiting endeavors. Meta-analysis studies should be conducted that expose the true utility of preemployment polygraph screening.

Research supports that age, prior military experience, and level of education have, to some degree, been shown as viable factors that predict cadet and future police officer performance (Aamodt, 2004; Peterson, 2002). Research also documents that GMA better predicts training and job performance than any other measure (Schmidt & Hunter, 1998), however, the current study did not support GMA (defined by test-taking ability) in predicting cadet training performance. Taking these ideas into context, it might be postulated that external stakeholders (society) would be better served by an officer who fits a certain demographic profile (i.e. age, military experience, level of education).

More effort is needed in the areas of research dedicated to general intelligence and polygraph results as they pertain to cadet performance. Empirical research documents the increasing utilization of polygraph testing in these settings. However, research fails to capture the essence of why government agencies place such trust in an instrument that is continually scrutinized for its controversy. Longitudinal studies that examine multiple polygraph testing techniques are needed to formulate hypotheses that either support or refute its usefulness in screening for the best cadet. Conducting research in these areas may tie together mechanisms that better predict cadet attrition and training performance; thus producing a better police officer.

Police agencies continue to evolve and scholars must produce research that is fruitful in the area of police personnel selection. More research is needed in the areas of effective measures of GMA and polygraph results as they relate to police personnel selection.

By examining current hiring procedures, police agencies stand a better chance to effectively implement strategies that compliment both agency and societal goals. Ultimately, this means incorporating hiring standards that are not only fair and legal, but remain competitive in order to facilitate the best possible outcome. From a police organizational perspective, a positive outcome is identifying hiring procedures that effectively capture the essence of what society demands; a competent officer that can protect his or her citizens. This, in turn, might provide police executives with insight as to the qualities they desire in a future officer. From a public safety standpoint, society is the benefactor by having the best qualified officers protecting their communities in an ever-changing world.

References

Aamodt, M. G. (2004). *Law enforcement selection*: *Research summaries*. Washington, DC.: Police Executive Summaries.

Aamodt, M. G., & Flink, W. (2001). Relationship between educational level and cadet performance in a police academy. *Applied HRM Research,* (6), *1*, 75-76.

Armstrong, D., & Polk, O. E. (2002). In W. W. Bennett, & K. M. Hess, (2004). *Management and Supervision in Law Enforcement* (4th ed.). Thompson Wadsworth: Belmont, CA.

Baker, S. A. (1995). Police academy training: Are we teaching recruits what they need to know? *Policing*, (21), *1, 54-79*.

Bennett, W. W., & Hess, K.M. (2004). *Management and Supervision in Law Enforcement* (4th ed.). Belmont, CA: Thompson Wadsworth.

Ben-Shakhar, G., & Furedy, J. J. (1990). *Theories and applications in the detection of deception.* A *psychophysiological and international perspective* (pp. 146-169) New York: Springer-Verlag.

Bernstein, I. H., Schoenfeld, L. S., & Costello, R. M. (1982). Truncated component regression, multicollinearity and the MMPI's use in a police officer selection setting. *Multivariate Behavioral Research,* (17) *1,* 99-116.

Bittner, E. (1990). *Aspects of Police Work.* Northeastern University Press, Boston: MA.

Brand, J., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, (75), *2*, 273-302.

Buerger, M. (2004). Educating and training the future police officer. *FBI Law Enforcement Bulletin*, (73), *1*, 26.

Campbell, D. J., & Campbell, K. M. (2010). Soldiers as police officers/Police officers as soldiers: Role evolution and revolution in the United States. *Armed Forces and Society*, (36), *2*, 327-350.

Capsambelis, C. (2004). Effective recruitment to attract the ideal law enforcement officer candidate. *Sheriff,* (56), *4*, 34-71.

Carter, D. L., Sapp, A.D. & Stephens, D. W. (1989). Police academy training: Are we teaching recruits what they need to know? *Policing*, (1), *21*: 54.

Chappell, T. A. (2008). Police academy training: Comparing across curricula. *Policing*, (31), *1*, 36-56.

Clay, O. J., Edwards, J. D., Ross, L. A., Okonkwo, O., Wadley, V. G., Roth, D. L., & Ball, K. K. (2009). Visual function and cognitive speed of processing mediate age-related decline in memory span and fluid intelligence. *Journal of Aging and Health*, (21), *4*, 547-566.

Cowper, T. J. (2000). The myth of the "military model" of leadership in law enforcement. Police Quartely, (3), 3, 228-246.

Decicco, D. (2000). Police officer candidate assessment and selection. *FBI Law Enforcement Bulletin*. (69), *12*, 1-6.

Fogleson, R. M. (1977). Predicting the effects of military service experience on stressful occupational events in police officers. *Policing*: *An International Journal of Police Strategies & Management*, (25), *3*, 602-618.

Forero, C. G., Pujol, D. G., Olivares, A. M., & Pueyo, A. A. (2009). A longitudinal model for predicting performance of police officers using personality and behavioral data. *Criminal Justice and Behavior*, (36), 6, 591-606.

Fowler, F. J. (2002). In Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods Approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Franz, V., & Jones, D. M. (1987). Predicting the effects of military service experience on stressful occupational events in police officers. *Policing*: *An International Journal of Police Strategies & Management*, (25), *3*, 602-618.

Fuss, T., & Snowden, L. (2004). Importance of background investigations. *Law & Order*, (52), 3, 58-63.

Gellner, U.B., Schneider, M.R., & Veen, S. (2011). Effect of workforce age on quantitative and qualitative organizational performance: Conceptual framework and case study evidence. *Organizational Studie*s, (32), *8*, 1103-1121.

Gottfredson, L. S. (1996). The triumph of techniques over purpose revisited: Evaluating police officer selection. *Review of Public Personnel Administration,* (21), *3*: 219-236.

Guller, M. (2003). *Predicting Performance of Law Enforcement Personnel Using the Candidate and Officer Personnel Survey and Other Psychological Measures*. (Doctoral dissertation), South Orange, NJ: Seton Hall University.

Guthrie, E. (2000). Higher learning and police training. *Law & Order*, (48), *12*, p. 124.

Hancock, B. W., & McClung, C. (1987). Psychological testing and the selection of police officers: A national survey. *Criminal Justice and Behavior*, (30), *5*, 511-537.

Harris, M., Dworkin, J., Park, J. (1990). Preemployment screening procedures: How human resource managers perceive them. *Journal of Business and Psychology*, (4) *3*, 279-292.

Henson, B., Reyns, B. W., Klahm, C. F. IV., & Frank, J. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. *Police Quarterly,* (13), *1*, 5-26.

Ho, T. (2001). The interrelationships of psychological testing, psychologist's recommendations, and police departments' recruitment decisions. *Police Quarterly*, (4), *3*, 318-342.

Hughes, F., & Andre, L. (2007). Problem officer variables and early-warning systems. *The Police Chief*, (74), 10. Alexandria, Virginia: International Association of Chiefs of Police.

Hunter, J. E., Schmidt, F. L. (1986). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, (124) *2*, 262-274.

Krapohl, D. (2002). In Kleiner, M. (2002). *Handbook of Polygraph Testing*. Academic Press: London, UK.

Jacobs, R., Solomon, T. (1977). Strategies for enhancing the prediction of job performance from job satisfaction. *Journal of Applied Psychology*, (62), *4*, 417- 421.

Jefferson, T. (1987; 1993). Predicting the effects of military service experience on stressful occupational events in police officers. *Policing*: *An International Journal of Police Strategies & Management*, (25), *3*, 602-618.

LaRose, A. P., Caldero, M. A., & Gutierrez, I. G. (2006). Individual values of Mexico's new centurions: Will police recruits implement community-based changes? *Journal of Contemporary Criminal Justice, (*22), *4*: 286-302.

Lewis, T.D., Shimerda, T.A. & Graham, G. (1983). What the academic advisor needs to know about job placement. *Journal of Accounting Education*, (1), 2, 135-142.

Lykken, D. T. (1981). *A Tremor in the Blood: Uses and abuses of the lie detector*. New York: McGraw-Hill.

Luszcz, M. A., & Bryan, J. (1999). Visual function and cognitive speed of processing mediate age-related decline in memory span and fluid intelligence. *Journal of Aging and Health*, (21), *4*: 547-566.

Meesig, R., & Horvath, F. (1995). National survey of practices, policies and evaluative comments on the use of preemployment polygraph screening in police agencies in the United States. *Polygraph*, (24), *2*, 57-136.

Ones, D., Viswesvaran, C., Schmidt, F. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, (78) *4*, 679-703.

Osoian, C., Zaharie, M., & Lazar, I. (2011). Does ownership matter? Employee selection practices in private and public sectors. *Transylvanian Review of Administrative Sciences*, (7), 33, 218-232.

Paoline, III Eugene A., Terrill, W. (2007). Police education, experience, and the use of force. *Criminal Justice and Behavior*, (34), *2*, 179-196.

Patterson, G. (2002). Predicting the effects of military service experience on stressful occupational events in police officers. *Policing*: *An International Journal of Police Strategies & Management*, (25), *3*, 602-618.

Randhawa, G. (2007). Work performance and its correlates: An empirical study. *Vision: The Journal of Business Perspective*. (11), *1*: 47-55.

Ree, M. J., J. A. Earles., & M. Teachout, M. S. (1994). What matters most? The perceived *importance* of ability and personality for hiring decisions. *Cornell Hospitality Quarterly, (52),* 2, 94-101.

Sanders, B. A. (2008). Using personality traits to predict police officer performance. *Policing: An International Journal of Police Strategies & Management,* (31), *1*, 129-147.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, (124), *2*, 262-274.

Schmidt, F. L. & Hunter, J.E. (2004). General Mental Ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, (86), *1*, 162-173.

Schroeder, D. J. (1973). *A study of the validity of the entrance examination for the position of patrolman under the guidelines established the Equal Opportunity Employment Commission.* (Unpublished master's thesis) New York, NY: John Jay College.

Sharit, J., Czaja, S. J., Hernandez, M., & Yang, Y. (2004). Investigating issues in aging and work performance using a customer service task simulator. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, (*48), *2*, 238-242.

Society for Industrial and Organizational Psychology (2003). Principles for the Validation and use of Personnel Selection Procedures. Retrieved from www.siop.org

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson.

Terman, L., & Otis, A. (1917). Predictive Validity of the MMPI-2 Psy-5 Scales and facets for law enforcement officer employment outcomes. *Criminal Justice and Behavior*, (37), *2*, 217-238.

Texas Department of Public Safety (2012). Retrieved from www.txdps.state.tx.us.

Texas Commisson on Law Enforcement Standards and Education (1997). Texas Peace Officer Job Task Analysis Report. Retrieved from http://www.tcleose.state.tx.us/

Texas Commisson on Law Enforcement Standards and Education (2008). Sunset Advisory Committee Staff Report. Retrieved from http://www.tcleose.state.tx.us/

Texas Commisson on Law Enforcement Standards and Education (2010). Basic Licensing Exam Results First Attempt Pass Rate 3 Year Summary (FY07 – FY09). Retrieved from http://www.tcleose.state.tx.us/

Texas Commisson on Law Enforcement Standards and Education (2012). Retrieved from h t t p : / / www.tcleose.state.tx.us/

Texas Department of Public Safety (2012). Retrieved from www.txdps.state.tx.us.

Trautman, N. E. (1986). Police academy training: Are we teaching recruits what they need to know? *Policing*, (21), *1*: 54.

Vodicka, A.T. (1994). Police academy training: Are we teaching recruits what they need to know? *Policing*, (21), *1*: 54.

Waddington, P.A.J. (1993). Predicting the effects of military service experience on stressful occupational events in police officers. *Policing: An International Journal of Police Strategies & Management,* (25), *3,* 602-618.

Walker, R. B. (1994). Police academy training: Are we teaching recruits what they need to know? *Policing*, (21), *1*: 54.

Waugh, L. (1996). *Police Recruit Selection-Predictors of Academy Performance*. City, State: The Criminal Justice Commission.

White, M. D. (2008). Identifying good cops early: Predicting recruit performance in the academy. *Police Quarterly*, (11), *1*, 27-49.

White, M.D. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. *Police Quarterly*, (13), *1*, 5-26.

Bonferroni and Šidák Corrections for Multiplicity Effects with Subtotal Scores of Comparison Question Polygraph Tests

Raymond Nelson

Abstract

The problem of multiple statistical comparisons is discussed as it applies to the use of subtotal scores of comparison question polygraph tests. Multiplicity phenomena, including inflation of alpha when any of a set of multiple subtotal scores are used to make deceptive classifications, and deflation of alpha when all of a set of multiple subtotals are used to make truthful classifications of test results. Common statistical corrections, including the Bonferroni correction and Šidák correction are described. Mathematical examples are provided to illustrate the application of these statistical corrections to the comparison question polygraph test.

Keywords: *scoring, test data analysis, polygraph techniques, normative data, reference distributions, statistical significance, Bonferroni, Šidák, multiple comparisons, inflated alpha, deflated-alpha, multiplicity*

Introduction

Multiplicity effects, also known as the problem of multiple comparisons (McDonald, 1996; Miller, 1981), are well known to scientists, researchers, statisticians and other professionals whose work involves the evaluation of data as a basis for classification and inference. These effects have also been referred to as the "look elsewhere" effect (White, 2011), because of the impulse or desire to continue to look elsewhere when we do not initially find what we are looking for. In the context of scientific research and testing, multiplicity effects, and the impulse to keep looking elsewhere until we find what we are looking for, can be thought of as a manifestation of a confirmation bias described by Nickerson (1998). We ignore results when we are unsatisfied and continue searching until we find a result with which we are satisfied.

A card-playing analogy can be useful to better understand the practical implications: image a poker player who deals himself a hand of cards with the goal of doing so repeatedly until he gets a Royal Flush. Probability theory tells us that with a sufficient number of trials, the odds will accumulate to a sufficiently high level that we are likely to eventually observe its occurrence. But, assuming a fair and unbiased deck of cards, it will be a mistake to attempt to infer that the deck of cards has any special characteristics or that the player has any unique attributes that caused the Royal Flush to occur. Instead, the occurrence of the Royal Flush is simply a function of continuing to look elsewhere (in subsequent hands of cards) for its occurrence. Similarly, looking repeatedly at any scientific dataset can confound our attempts to make realistic and accurate inferences about significance or meaning when we eventually observe what we are looking for. More specifically, multiplicity effects are the compounding of error probabilities. They can result in a loss of accuracy or precision and corresponding increase in classification error.

Discussion

Multiplicity effects play a role in comparison question polygraph examinations when using subtotal scores to classify the results as deceptive or truthful. Subtotal scores for individual relevant questions have been shown to be an effective basis for deceptive classifications when the grand total score is inconclusive (Senter, 2002; Senter & Dollins, 2003; 2008). But polygraph techniques that make use of grand total scores have consistently produced higher accuracy rates than techniques for which decisions are based solely on subtotal scores (APA, 2011). Subtotal scores have been the traditional basis with which to classify the results of multiple issue screening exams (Department of Defense, 2006a, 2006b) when hand scoring. The grand total hand scores are not traditionally used in multiple issue screening tests.

Use of polygraph subtotal scores as a

basis for statistical classification and inference will introduce known and predictable mathematical and statistical increases to the probability of error unless corrections are applied. These effects occur because every imperfect and non-deterministic test result is a probabilistic result. There is always some associated probability that the result is correct and incorrect. The test error estimate will be an aggregation of the errors for all probabilistic results used to classify the test result.

Inflation of alpha for deceptive results of event-specific diagnostic polygraphs.

Use of subtotal scores in event-specific examinations, for which one classification will be made at the level of the test as a whole, introduces multiplicity into the statistical model. It amounts to the practice of making multiple statistical decisions regarding a single classification. When making multiple probabilistic judgements regarding a single target incident or allegation, for which any deceptive subtotal result will result in the classification of the examination as deceptive, the resulting probability of error is the cumulative or additive probabilities of error for all subtotal probability scores. In the case of an event-specific diagnostic polygraph with three relevant questions (RQs) and alpha = .05, the total error probability can be determined by summing the alpha levels for all RQs (.05 + .05 + .05 = .15). Calculations indicate a potential for a 15% error rate even though the test is conducted with alpha = .05, with the goal of constraining errors to a rate less than 5%. This has sometimes been referred to as the problem of *inflated alpha* because of the predictable increase in test errors. Left unmanaged, inflated alpha can result in a false positive error rate that is potentially several times greater than that which was intended or anticipated. So while the goal was to constrain false positive errors to 5%, the practice of using subtotals increased that false positive error rate to about 15%.

Bonferroni correction.

Fortunately, the problem of inflated alpha is only mildly vexing and is quite easily rectified through the use of a simple statistical correction - the *Bonferroni correction* (Abdi, 2007), named for famous Italian statistician Carlo Emilio Bonferroni (1892-1960). The Bonferroni correction is calculated by dividing the desired alpha level by the number of statistical decisions. The number of statistical decisions is equal to the number of subtotal scores which is the same as the number of RQs. The resulting corrected alpha level is referred to as the *Bonferroni corrected alpha*.

For event-specific diagnostic polygraph with three RQs and a desired alpha of .05 we divide .05 by three (alpha = .05 / 3 RQs = alpha .0167 per RQ). It will be necessary to use the Bonferroni corrected alpha = .0167 for each of the three subtotal scores. When these per question error probabilities accumulate (.0167 + .0167 + .0167 = .05) the total cumulative margin of error for the test will be alpha = .05. The error estimate will be constrained to within the desired range of less than 5%. Event-specific diagnostic exams with two RQs will require the use of Bonferroni corrected alpha = .025. This is because alpha = .05 / 2 RQs = .025 per RQ, and this will accumulate to .025 + .025 = .05. Similarly, event-specific diagnostic exams with four RQs will use Bonferroni corrected alpha = .05 / 4 = .0125 per RQ, which will accumulate to .0125 + .0125 + .0125 + .0125 = .05. Because subtotal scores are not used to make truthful classifications for event-specific diagnostic exams, no statistical correction is needed for truthful classification for these type of exams.

Deflation of alpha for truthful results of multiple-issue screening exams.

Multiple-issue screening exams make use of subtotal scores for both deceptive and truthful classifications. This is accomplished with the "any or all" rubric which states that *any* subtotal result that is deceptive will be sufficient to classify the exam result as deceptive, whereas *all* subtotal results must indicate truth-telling in order to classify the overall exam result as truthful. Results are inconclusive whenever one or more of the subtotal scores are not statistically significant for truth-telling and none of the subtotal scores is statistically significant for deception. As with event-specific diagnostic exams the test error statistic for a multiple-issue screening polygraph is a function of the number of subtotal scores (the number of relevant questions). This phenomena applies to all forms of testing that involve multiple statistical comparisons.

With event-specific diagnostic exams all relevant questions describe details related to a single allegation or incident. Relevant questions for multiple-issue screening polygraphs will describe different behavioral issues with a strong assumption of independence. The independence assumption is not premised

solely on the use of different action verbs or semantic content for each relevant question. It also involves the assumption that an examinee could engage in one or more behaviors while conceivably remaining completely uninvolved in other behaviors. (This independence assumption is not used with event-specific diagnostic polygraph for which the all relevant questions describe aspects of a single allegation or incident.)

The independence assumption is said to be a strong assumption because, in reality, although the target *behaviors* might be assumed to be independent or unaffected by one another, even though the examinee´s *responses* to multiple issue polygraph stimulus questions are *not* completely independent. Responses to different target stimuli can affect one another within an exam. This is because all responses to multi-issue screening stimuli have an important source of shared variance - the examinee. The fact that responses are not completely independent appears to be the basis of the need for the "any or all" rubric and for traditional prohibitions against attempting to make both deceptive and truthful classifications within a single examination. Because the any or all rubric does not allow both truthful and deceptive results, it will eliminate the potential to observe both false positive and false negative errors within a single examination. Instead, observed testing errors will be in the form of either false-positive or false-negative errors, for which we can constrain their occurrence to desired levels.

Because the target issues for multiple-issue screening exams are treated independently, there is no great concern that we are subjecting a single target issue to multiple statistical decisions. Screening tests are intended to identify possible problems that can be subsequently evaluated in more thorough detail, and test sensitivity is therefore an important concern. Statistical corrections are not used when there is a potentially costly loss of sensitivity that would reduce the test effectiveness (McDonald, 2009). For these reasons, Bonferroni correction is not used to make deceptive classifications for multiple-issue screening polygraphs. Deceptive classifications of multiple-issue screening polygraphs are made with the uncorrected alpha boundary.

Multiplicity plays an important role in *truthful* classifications for multiple-issue screening polygraphs, but in a slightly different way. Truthful classifications are made when the observed data differ at a statistically significant level from the statistical reference distributions for deceptive cases. Alpha for truthful classifications therefore represents the tolerance for risk or error that a deceptive person may be classified as truthful in a multiple issue screening test (a false-negative error). Quite obviously, most deceptive persons can be expected to produce deceptive test scores, and the proportion of deceptive persons that produce a test question score that is statistically significant for truth-telling (i.e., differs at a statistically significant level from the normative reference distributions for deceptive cases) is expected to be observed at the defined alpha level (.05). Perhaps equally obvious is the fact that the proportion of deceptive persons that produce *two* statistically significant truthful scores in a test with two relevant questions will be lower than the proportion of deceptive persons who produce only one statistically significant truthful score. Similarly, the proportion of deceptive persons who produce three out of three truthful scores, or four out of four truthful scores, can be expected to be even lower. This phenomena can be thought of as the *deflation of alpha* that occurs as a result of the requirement that the examinee pass *all* questions in order to pass the test. Deflation of alpha will result in a reduction of the observed false-negative error rate to something predictably lower than the established alpha tolerance for error.

Deflation of alpha will reduce testing errors for deceptive classifications, but will also have an effect on truthful classifications. The requirement that all subtotal scores are statistically significant for truth-telling will effectively provide the truthful examinee with multiple opportunities to not produce a statistically significant truthful score. This is a simple feature of the fact that all tests are probabilistic and not deterministic, and that probabilities can be cumulative under these circumstances. For this reason, the requirement for statistically significant truthful scores for *all* subtotals can be expected to cause a substantial *inflation of inconclusive* results for truthful persons, along with a corresponding substantial reduction of test specificity for truth-telling - unless a statistical correction is used.

Šidák correction.

The preferred statistical correction for truthful classifications of multiple-issue screening polygraphs is not the Bonferroni

correction but is instead a related procedure called the *Šidák correction* (Abdi, 2007, Šidák, 1967). The Šidák correction is named for Zbyněk Šidák (1933-1999), a renowned Czech statistician. It is an exact version of the simple Bonferroni correction that is better suited to the context of multiple independent classifications. Calculation of the Šidák correction is thus: $1-(1-\text{alpha})^{\text{number-of-decisions}}$. The Šidák correction is the mathematical compliment of the compliment of the alpha raised to number of decisions. As with the previously described Bonferroni correction, the number of decisions is equal to the number of subtotal scores which is also equal to the number of relevant questions.

The normal form of the Šidák correction is used to calculate the inflation of alpha. But we are concerned with the deflation of alpha, so it will be the inverse of the Šidák correction that is used calculate this deflation. The inverse of the Šidák correction is calculated using the following equation: $1-(1-\text{alpha})^{1/\text{number-of-decisions}}$. The inverse Šidák is the mathematical compliment of the compliment of the alpha raised to the inverse of the number of decisions.

To demonstrate the application of the Šidák correction to adjust or correct the alpha boundary for the number of relevant questions, consider the following example: a multiple issue polygraph with 4 relevant questions for which alpha = .05 will give the following uncorrected, deflated, alpha level: $1-(1-.05)^{1/4} = .0127$. That means that instead of constraining false negatives to our desired 5% we actually constrain them to 1.27%. This will result in a corresponding increase in inconclusive truthful cases. Correcting for this will involve first calculating the corrected alpha boundary using the normal form: $1-(1-.05)^4 = .1854$. Use of the Šidák corrected alpha = .1854, will give the following: $1-(1-.1854)^{1/4} = .05$. This will preserve the test specificity to truth-telling for multiple-issue screening polygraphs at acceptably high levels, while reducing the occurrence of inconclusive results for truthful persons. It will also constrain the occurrence of false-negative test errors to rates that are within the tolerance level expressed by the alpha = .05 level.

In practice, statistical corrections can be applied to either the alpha boundary or to the p-values using either the normal or inverse forms. However, correction of p-values can only be accomplished after conducting and scoring an examination, whereas correction of alpha boundaries can be accomplished prior to the conduct of an examination. This is accomplished by using the subtotal score with a p-value at or below the corrected alpha as a decision threshold.

Conclusion

All scientific testing is a process of classification and inference. Classification, in this case, refers to the formulation of a simple categorical test result. Inference is the process of calculating a statistical or probabilistic estimate of the likelihood that an error has occurred. In a more abstract sense, the purpose of scientific testing is to evaluate and quantify an amorphous phenomena that cannot be subjected to simple and perfect deterministic observation or to direct physical/linear measurement. Deterministic observation requires the existence of some phenomena that is uniquely and perfectly associated with the thing we want to evaluate. This would be theoretically perfect, and would also obviate the need for testing. Physical measurement, in contrast, is near perfect, though still subject to mechanical measurement error, and would require two things: 1) a physical substance to measure, and 2) a well-defined unit of measurement. Scientific tests are inherently probabilistic - they are neither deterministic nor an actual physical measurement. Scientific tests are not expected to be perfect. They are expected to quantify the probabilistic margin of uncertainty surrounding a conclusion. Good scientific tests will do this in manner such that the predicted proportions of testing errors concurs reasonably with the observed evidence of testing errors. Multiplicity effects have a potentially serious impact on the accuracy of test error estimates. The use of statistical corrections can be an important part of the validity and effectiveness of a test method.

Two core ideas underlie all scientific tests and experiments. The first core idea is that all scientific conclusions or hypotheses are relative to some alternative. Professionals who make scientific conclusions are expected to articulate the alternatives and to use probability theory to weigh the evidence. The second core idea is that all conclusions and hypotheses must be stated as statistical or probabilistic hypotheses in order to be quantifiable. Conclusions or hypotheses that cannot be stated as statistical hypotheses cannot be measured or tested, and are therefore not sci-

entific. Unscientific ideas that portend to be scientific can be said to be pseudoscience.

Related to the need for testable hypotheses is the need to make *a priori* declarations about the tolerance for error and required alpha level for statistical significance. Field practitioners generally do not themselves decide on alpha boundaries or numerical cutscores - these are most often a matter of agency policy and are developed around the needs specific to the risk management context. Field practitioners themselves are also not expected to calculate statistical formulae themselves. Instead, they commonly use published statistical reference tables for which calculations have been previously computed for all possible test results.

If the polygraph test is merely a tool to amplify or enhance an interrogation or interview, then examiners need not ever account for or explain the test results. If this were the case they need not even score the test, and certainly need not learn about probability theory and statistical phenomena. Similarly, polygraph examiners will never be expected to account for or explain a test result if a confession is obtained for every deceptive test result without fail. If the information from the pretest and posttest discussions are the sole purpose of the polygraph test then there would be no need to ever provide a test result. If, however, there is ever a need to explain a test result or account for the level of certainty or uncertainty that should be attributed to a test result, examiners might be obligated to numerically score and statistically quantify the test result. Examiners who are unprepared to do this will be vulnerable to professional embarrassment, either due to an inability to provide evidence based computations of the expected test precision and error rate, or due to frustration when it is eventually discovered that polygraph results are probabilistic and imperfect despite a feigned attitude of certainty.

Examiners who are prepared to account for test results using the basic principles and concepts of statistics and probability and theory will be better prepared to make favorable professional impressions while discussing test results without the sense of insecurity that stems from naive expectations for deterministic perfection from a probabilistic test. Although there will always be practical value in in the information that can be obtained from the polygraph pretest and posttest interviews, test results without realistic computations of statistical error estimates will, in the end, be of no real value.

Ultimately, all test scores, including both grand total and subtotal scores of comparison question polygraph tests, will have an associated probability of error. Probability theory informs us that error rates are predictably cumulative whenever we attempt to make multiple statistical comparisons within a single test or experiment. While mildly concerning, the predictability of multiplicity phenomena means that we can also apply the principles of probability theory to statistically correct for multiplicity effects - if we understand the principles of probability. While very simple calculations such as the Bonferroni correction can be easily managed in field settings, field practitioners should be relieved of complex calculations such as the Šidák correction through the inclusion of statistically corrected information in published normative reference tables. Use of computer algorithms can also accomplish the application of these statistical corrections with automated reliability. Although many researchers, statisticians and scientists will prefer to use omnibus statistical methods such as ANOVA and other methods to simultaneously test multiple statistical hypothesis without the introduction of multiplicity effects, Bonferroni correction and Šidák correction are two classical solutions to the well know problems of multiplicity. They are well suited to the analysis and interpretation of comparison question polygraph test results.

References

Abdi, H. (2007). Bonferroni and Šidák corrections for multiple comparisons. In N.J. Salkind (Ed.), *Encyclopedia of Measurement and Statistics.* Sage.

Department of Defense (2006a). *Federal psychophysiological detection of deception examiner handbook.* Reprinted in Polygraph, 40 (1), 2-66.

Department of Defense (2006b). *Test data analysis: DoDPI numerical evaluation scoring system.* [Retrieved from http://www.antipolygraph.org on 3-31-2007].

McDonald, J. H. 2009. *Handbook of Biological Statistics 2nd ed.*. Sparky House Publishing, Baltimore, Maryland.

Miller, R. G. (1981). *Simultaneous Statistical Inference 2nd Ed.* Springer Verlag New York.

Nickerson, R. S. (June 1998). "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises". Review of General Psychology 2 (2): 175–220.

Šidàk, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association, 62,* 626–633.

White, L. A. (August 12, 2011). *Word of the Week: Look Elsewhere Effect.* Sanford National Accelerator Laboratory.

Letter to the Editor Regarding article by Nelson and Handler entitled Statistical Reference Distribution for Comparison Question Polygraphs.

James Allan Matte

Dear Editor:

This letter pertains to Appendix P, Matte Quadri-Track Zone Comparison Technique, of article entitled Statistical Reference Distributions for Comparison Question Polygraphs by Raymond Nelson and Mark Handler, *Polygraph*, Volume 44, Nr. 1, 2015.

In Footnote #9, Nelson and Handler, referring to the 2011 APA meta-analytic survey, stated "Studies supporting this technique have been described as substantially methodologically flawed, and it is considered unlikely that the reported accuracy rates will be achieved in field settings." The three field studies validating the Quadri-Track ZCT were in *field setting*s (Matte, Reuss 1989b; Mangan, Armitage, Adams 2008a; Shurany, Stein, Brand 2009), and the studies were not substantially flawed as indicated in this author's critique (Matte 2012). In fact, the aforesaid field studies met the most stringent requirements set forth in the *Guiding Principles and Benchmarks for the Conduct of Validity Studies of Psychophysiological Veracity Examinations Using the Polygraph* (Matte 2010), requiring a minimum sample of 50 confirmed cases (Matte 122, Mangan 140, Shurany 57). Conversely, the APA meta-analytic survey listed four studies that used sample cases from 20 to 30 cases validating their respective evidentiary techniques. One of them, the Nelson, Handler, Blalock, Cushman 2012 field study with a sample of 22 cases (Polygraph, In Press) has not been published as of 6 January 2015 (R. Nelson, personal communication 6 January 2015). Sample size has a direct relationship to the applicability of the study's results to the general population. As explained in detail in the aforementioned *Guiding Principles and Benchmarks*, several important elements present in field studies are lacking in laboratory studies, which is beyond the scope of this Letter to the Editor which APA now limits to 400 words, one table and 10 references.

In Footnote #9 Nelson, et al stated "published procedures for this technique involve the average total score per chart instead of the more common grand total score." This statement is inaccurate as reflected in diagram below and several published articles and studies listed in the *unabridged* 2000 word Letter-to-the-Editor published on website at www.mattepolygraph.com under heading of *Publications by James Allan Matte.*

The Quadri-Track ZCT Numerical Score Sheet and Conclusion Table

STIMULATION TEST DATA:	NUMBER SELECTED: CHART NUMBER:							

Quadri-Track Tri-Zone Quantification System Score Table

CHART 1	NDI	INDEF	DI			NDI	INDEF	DI			NDI	INDEF	DI		
PNE (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
EDA (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CAR (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CHART 2	NDI	INDEF	DI			NDI	INDEF	DI			NDI	INDEF	DI		
PNE (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
EDA (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CAR (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CHART 3	NDI	INDEF	DI			NDI	INDEF	DI			NDI	INDEF	DI		
PNE (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
EDA (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CAR (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CHART 4	NDI	INDEF	DI			NDI	INDEF	DI			NDI	INDEF	DI		
PNE (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
EDA (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
CAR (33)	+3 +2	+1 0 -1	-2 -3	= ( )	(35)	+3 +2	+1 0 -1	-2 -3	= ( )	(24)	+3 +2	+1 0 -1	-2 -3	= ( )	
TARGET ( )		TOTAL: ( )				TOTAL: ( )					TOTAL: ( )				

GRAND TOTAL: ( )
FOR ( ) CHARTS.

% Pop:_____
P.E.: _____

CONCLUSION TABLE

	CIRCLE APPROPRIATE NUMBER BELOW		
RESULTS FOR 1 CHART	+27 to +3	+2 to -4	-5 to -27
	TRUTH	INDEFINITE	DECEPTION
	CIRCLE APPROPRIATE NUMBER BELOW		
RESULTS FOR 2 CHARTS	+54 to +6	+5 to -9	-10 to -54
	TRUTH	INDEFINITE	DECEPTION
	CIRCLE APPROPRIATE NUMBER BELOW		
RESULTS FOR 3 CHARTS	+81 to +9	+8 to -14	-15 to -81
	TRUTH	INDEFINITE	DECEPTION
	CIRCLE APPROPRIATE NUMBER BELOW		
RESULTS FOR 4 CHARTS	+108 to +12	+13 to -19	-20 to -108
	TRUTH	INDEFINITE	DECEPTION

References

American Polygraph Association. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph*, 40(4): 193-305.

Mangan, D. J., Armitage, T. E., Adams, G. C. (2008a). A field study on the validity of the Quadri-Track Zone Comparison Technique. *Physiology & Behavior*, 95, 17-23.

Matte, J. A. (2012). Critique of Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph Techniques. *European Polygraph*, 6, 1(19): 19-44.

Matte, J. A., Reuss, R. M. (1989b). A Field Validation Study on the Quadri-Zone Comparison Technique. *Polygraph*, 18(4): 187-202.

Nelson, R. Handler, M. (2015). Statistical Reference Distributions for Comparison Question Polygraph. *Polygraph*, 44(1): 91-114.

Response to James Allan Matte Letter to the Editor

Raymond Nelson

Appendix P in Nelson and Handler (2015) is calculated from the statistics published on page 98 in Matte and Reuss (1989), which recommends cut-scores of -5 and +3 per chart and includes an instruction to average the scores for all charts. We know of no peer-reviewed publication that recommends the use of any other distribution or cutscores. Mr. Matte´s suggested cutscores for 2, 3 and 4 charts are simply multiples of these cutscores - disregarding that standard deviations are not subject to linear addition and multiplication. Instead of attempting to rectify Mr. Matte´s statistical and scientific inconsistencies we elected to republish his recommendations. Our concerns about methodological issues were included in footnotes.

Mr. Matte´s citation of a self-publication cannot be taken to re-define how scientific research and statistical analysis actually work. In particular, Mr. Matte´s assertions about sampling and generalizability are wrong. Sample size does *not* affect the generalizability of scientific conclusions; sampling method does. Sample size *does* affect statistical power – the ability to find a significant effect – and this will be important when investigating small effect sizes. Polygraph research is commonly seeking large effect sizes - large improvements over chance - for which smaller sampling sizes are often adequate. It is not surprising that Mr. Matte´s research sample, consisting of examinations conducted or supervised by himself (using his eponymously named technique), included no error cases and ~100% classification accuracy. We hypothesize that reliance on confessions as a sampling method may have systematically excluded both false-positive and false-negative error cases, for which a confession is not likely to be obtained, from the reported study samples.

If Mr. Matte is correct in his conclusion of ~100% accuracy, then the problem of perfect lie detection has been solved. No further research is necessary, and there is nothing more we need to learn. If Mr. Matte is incorrect, if ~100% polygraph accuracy cannot be achieved by most examiners, most of the time, with most examinees, then Mr. Matte´s conclusions would appear to tell us little, if anything, about what to expect in reality.

We thank Mr. Matte for letting us know his feelings, but we remain in disagreement with his assertions and conclusions. We apologize for subjecting the readership to another round of argument and controversy in a matter for which there will be no benefit to the profession.

Sexual History Disclosure and Sex Offender Recidivism

James Edward Konopasek, Ph.D.

Colorado State University – Global Campus

Raymond Nelson, MA, NCC

Private Practice

Abstract

This research examines the extent to which non-deceptive sexual history polygraph test results are associated with treatment outcomes and sexual recidivism. Specifically, the project examined correlations among several independent variables (i.e., sexual history polygraph results, risk level/score, age at which a non-deceptive sexual history polygraph was achieved, achievement of a non-deceptive sexual history polygraph result within six months of treatment onset, sexual deviance, psychopathy, and denial) and the dependent variables of treatment completion status and sexual recidivism. A cohort of 170 convicted sexual offenders was evaluated for a period of five years following completion of treatment or discharge from supervision. Analysis revealed that the achievement of a non-deceptive sexual history polygraph result was moderately associated with completion of treatment ($r\phi = .328$, $p < .001$). Two variables, achievement of a non-deceptive sexual history polygraph result within six months of treatment onset ($r\phi = -.152$, $p = .047$), and age under 35 at the time of a non-deceptive sexual history polygraph ($r\phi = .167$, $p = .029$) were shown to be correlated with sexual recidivism. This research provides preliminary evidence that non-deceptive sexual history polygraph results are associated with favorable treatment and recidivism outcomes.

Keywords*: sex offender, rehabilitation, recidivism, Static-99R, polygraph, disclosure*

Sexual History Disclosure and Sex Offender Recidivism

Sexual history polygraph testing, is an adjunct component of approximately two thirds (67%) of the adult outpatient sex offender treatment programs in the U.S. (McGrath, Cumming, Burchard, Zeoli & Ellerby, 2010). Sexual history polygraph examinations are conducted in an effort to motivate full, accurate, and timely disclosure of sexually deviant behavior (Ahlmeyer, Heil, McKee & English, 2000; Association for the Treatment of Sexual Abusers, 2004; Emerick & Dutton, 1993; O'Connell, 1998). Polygraph testing is intended to encourage offenders to fully disclose their history of sexual offending behaviors. Treatment providers who make use of sexual history polygraph testing do so with the goal of the identification of paraphilic and crossover offense behaviors so they can formulate accurate and effective treatment planning to facilitate the cessation of continued offenses (Ahlmeyer, Heil, McKee & English, 2000; Bourke & Hernandez, 2009; O'Connell, 1998; Wilcox & Sosnowski, 2005). For more information on the use of polygraph testing in sex offender management see (English, 1998; English, Jones, Pasini-Hill, Patrick & Cooley-Towell, 2000; Grubin, 2008; Hindman & Peters, 2001; Levenson, 2009). Skeptics of polygraph testing have pointed to unanswered questions regarding the contribution of polygraph testing to observable and measurable outcomes (Ben-Shakhar, 2008; National Research Council, 2003; Rosky, 2013). Despite the existence of controversy, polygraph testing has become a recognizable component of sex offender supervision and treatment programs.

Author's Note
Correspondence Address: PO Box 178, Felton, PA 17322
Email: james.konopasek@csuglobal.edu

The Importance of Timely, Honest, Disclosure

Farber (2003) and Farber and Hall (2003) discussed the importance of disclosure in psychotherapy and how this process includes inherent challenges. Obstacles to timely, honest, disclosure of sexual problems include feelings of shame, guilt and fear among clients--which may contribute to deliberate secret-keeping, things left unsaid in therapy sessions, and what is termed *the untold story* relating to clinically relevant history. With general psychotherapy patients, the following have been described as the most prevalent items not disclosed: sexual and body-oriented experiences, sexual feelings and fantasies toward the therapist, interest in pornography, bathroom habits, experiences and feelings toward masturbation, loss of virginity and fidelity (Farber & Hall, 2002). Nearly all of these topics will also apply to the treatment of sex offenders.

Recent research in sex offender treatment has shown that measurement of the quantity and seriousness of clinically relevant disclosures (CRDs), obtained through polygraph testing, may be useful to identify and adjust treatment targets (Gannon, Wood, Pina, Tyler, Baroux & Vasquez, 2014). Gannon, et al., quantified CRDs in four categories: thoughts, feelings and attitudes (e.g., abusive fantasies and desires); sexual behavior (e.g., use of pornography); historical information (e.g., admitting unknown offense behavior); and changes in circumstance/risk-behaviors (e.g., increased access to children). They reported that sex offenders who were subject to polygraph testing made CRDs in 572 sessions versus 320 CRDs for controls not subject to polygraph testing, but also found that the seriousness ratings of disclosures did not differ across the two groups. Gannon et al. cautiously described the evidence showing that user satisfaction benefits were expressed by treatment and supervision professionals who made use of maintenance/compliance testing with sex offenders in the United Kingdom.

Although less is known, at the present time, about the benefits of the sexual history polygraph, there is general and emerging evidence that timely self-disclosure of problem behaviors is viewed by professionals as a favorable indicator that may be a factor in effective risk assessment, treatment planning and supervision in the community. How disclosure is measured varies, and the relevance of disclosure to treatment and recidivism outcomes remains elusive. This research aims to fill some of this gap in knowledge.

Accuracy and Validity of Sexual History Polygraph Testing

Polygraph examinations have been used to measure the veracity of sex offender disclosure for over 30 years (Abrams, 1991), beginning in the early 1980s. The accuracy of such testing has been questioned. Despite continued controversy, there is a substantial and growing literature supporting the polygraph as capable of discriminating deception and truth-telling at rates significantly greater than chance; and publishing the validity and reliability of various polygraph testing techniques, with both favorable and unfavorable findings (Furedy, 1996; Honts, 1996; Honts & Alloway, 2007; Iacano, 2008; Iacano & Lykken, 1997; Krapohl, 2006; National Research Council, 2003; Offe & Offe, 2007; Nelson, Handler, Krapohl, Gougler, Shaw & Bierman, 2011). Research and publication has also addressed threats to polygraph accuracy (Honts, Hodes & Raskin, 1985; Honts, Raskin & Kircher, 1994; National Research Council, 2003; Patrick & Iacano, 1989).

Most relevant to this study is research on the accuracy of two types of Comparison Question Technique (CQT) exams (i.e., the Air Force Modified General Question Technique – AFMGQT, and the Directed Lie Screening Test - DLST). The AFMGQT and DLST are considered by polygraph examiners to be equally well suited for multiple issue screening polygraphs in the post-conviction and other settings. Both techniques involve the same instrumentation and sensors. Differences between the two formats primarily involve the procedural rules for recording of several presentations of the test stimulus questions.

The scientific basis of the CQT is that difference in the pattern of response to relevant and comparison test stimuli can be observed as a function of deception or truth-telling in response to the relevant stimuli. Used in the context of a sexual history polygraph, the premise of the CQT is that an examinee who withholds information about his or her sexual history will produce physiological reactions that are loaded onto sexual history questions, whereas an examinee who is truthful will produce responses that are loaded on comparison stimuli. Senter, Weatherman, Krapohl, and Horvath (2010) referred to this phenomenon as *differential salience*. Polygraph results can be analyzed for

their statistical significance, and can also be described categorically. Polygraph examiners have adopted the categorical terms *Significant Reaction* and *No Significant Reaction*, though some results are described using the more traditional categorical terms *Deception Indicated*, or *No Deception Indicated* that remain in more common use in diagnostic polygraph contexts. These categorical labels are the contextual analog for the more abstracted terms Positive and Negative as used in other scientific testing context.

During recent years, there has been an increase in published polygraph research on the validity of the wide variety of polygraph testing techniques (Handler, Nelson & Blalock, 2008; Krapohl, 2006; Nelson, et al., 2011). Nelson, et al., (2011) summarized the results of 14 studies involving 1,008 cases and 31 different scores, and described the accuracy of polygraph techniques scored with an assumption of independent criterion variance, such as those used in post-conviction sex offender testing, as proving a mean unweighted accuracy rate of .850, with a 95% confidence range from .773 to .926.

Polygraph testing, applied to the sexual history disclosure polygraph, will address whether an offender has truthfully disclosed the details of his or her history of sexual offense behaviors. Because there is no known allegation or incident that is the target of the exam or investigation, that exams are *screening tests,* also referred to as exploratory exams, and investigative techniques (Handler, Nelson & Blalock, 2008), and are not intended to become the sole or central basis for decision and action in the same way manner as a diagnostic testing context.

Relevant questions for sexual history screening polygraphs can address several types of sex offending behavior, including: sexual offenses against underage children since becoming an adult, sexual contact with relatives and family members, forced sexual contact/violent sexual offenses, and sexual contact with persons who were asleep or unconscious. Relevant questions can also address sexual behaviors that can signal problems involving sexual compulsivity or sexual preoccupation, such as voyeurism, exhibitionism, public masturbation, stalking behaviors, theft or use of underwear/ undergarments or personal property for masturbation or sexual arousal, child pornography, and other sexual behaviors that may indicate problems with sexual deviancy.

Sexual History Polygraph Testing in the Sex Offender Management Context

Several studies describe the increase in disclosure of information (Abrams, 1991; Ahlmeyer, et al., 2000; English et al., 2000; Grubin, Madsen, Parsons, Sosnowski & Warberg, 2004; Gannon, et al., 2014; Heil et al., 2003; Kokish, Levenson & Blasingame, 2005; O'Connell, 1998; Wilcox, Sosnowski & Middleton, 1999) that results from polygraph testing, and that may contribute to improved risk assessment, treatment planning, and case management. Fewer studies have addressed the relationship between polygraph testing and *recidivism*. McGrath, Cumming, Hoke and Bonn-Miller (2007) and Cook (2011) are the only studies we located that address this relationship.

McGrath, et al. (2007), using a comparative matched-pairs design, studied recidivism incidence among treated offenders who were subject to maintenance polygraph testing regarding compliance with the supervision and treatment program. One hundred four (104) polygraph cases were matched (relative to completion status, risk level and offense severity) with 104 treated sex offenders who did not undergo polygraph testing. Although no relationship to sexual recidivism was found, the polygraph group showed lower *violent recidivism* rates than the non-polygraph group, 2.9% versus 11.5%. Cook (2011) reported results using logistic regression and a modified version of the Static-99 risk assessment instrument that included additional variables gleaned from sexual history polygraph examination (SHPE) reports. Those results showed that data extracted from SHPE reports on 93 convicted sex offenders, including early onset of sex offending prior to age 13, more than 2 admitted paraphilias, and passing/failing the SHPE did not account for any statistically significant change in the odds of violent, including sexual, recidivism. The present study is similar to McGrath, et al. (2007) and Cook (2011) in that it is aimed at increasing the body of knowledge on polygraph-assisted interventions related to treatment outcome and sex offense recidivism.

Purpose and Method

The purpose of this study is to explore correlations among variables related to test results from sexual history polygraph testing, treatment outcome, and sexual recidivism among convicted sex offenders. This study also examines relationships between study

variables and several other literature-derived variables that have been shown to be correlated with sexual recidivism. Those variables include: age, sex offender type, Static-99R risk score/level, denial, sexual deviance, and psychopathy. Relationships among the study variables were analyzed using bivariate analyses utilizing the Phi Coefficient for dichotomous variables and Point Biserial Correlation.

Assumptions

A key assumption in this project is that the CQT polygraph screening examinations conducted on individuals in this sample are similar enough in design and administration, and criterion accuracy rates are similar to those described by Nelson et al. (2011), specifically the AFMGQT. Polygraph examination reports utilized in this study indicated that CQTs were administered in all cases and that relevant sexual history questions were constructed around the target issues described previously in this report. However, the exact nature of comparison questions, scoring rules, and decision cut-off scores are unknown.

Procedures

Data were extracted from hard-copy files stored in program archives and from the computerized database (Microsoft Access) of an outpatient treatment program. Consultation during the design phase of this project (O'Connell, 2011) resulted in the use of independent third-party research assistants to ensure that the principal author remained blind as to the identity of recidivist identity in order to alleviate the potential for bias or conflict of interest. A research assistant constructed three research data sets: 1) a small 8-item identified data set for the purposes of requesting and capturing criminal history recidivism data from criminal history gatekeeper agencies, 2) a 64-item *identified* data set containing all variable data, and 3) a 52-item *unidentified* data set. Only the *un*identified data set was shared with this researcher. The unidentified data set was exported to a SPSS database file for analysis.

Data was captured for the dependent variable, recidivism, by a research assistant who conducted public record criminal history searches in accordance with the gatekeeper agency's confidentiality, terms of use, storage, retention and destruction guidelines (LexisNexis Accurint, 2011; Oregon Judicial Department--Justice Information Network--OJIN, 2009; Washington State Patrol--WATCH, 2011; Washington State Institute for Public Policy, 2009).

Test results were extracted from archived polygraph examination reports and pretest questionnaires, along with information on whether a non-deceptive SHPE result was achieved within six months of treatment onset. These data were entered into the larger, 64-item, database. Criminal history recidivism data was then matched for each case by the research assistant. Archived paper records included psychosexual evaluations, presentence reports, polygraph examination reports and treatment progress reports. Static-99R risk computation forms were completed by the principal author, who received training from the Justice Institute of British Columbia (JIBC, 2011) on the scoring of the Static-99R in accordance with published coding rules (Harris, Phenix, Hanson & Thornton, 2003; Helmus, Babchishin, Hanson, & Thornton, 2009) during 2011. These data were entered into the identified database, after which the research assistant entered matching recidivism data into the dataset. Due to time, personnel and training budget constraints of this self-funded project, third party Static-99R scorers were not used, and inter-rater reliability statistics were not obtained.

Sample

The study involved a convenience sample of adult male sex offenders (N = 170) mandated by the courts and parole/probation agencies to be evaluated and treated at four outpatient sex offender treatment programs between 1994 and year-end 2004. The sample is a small fraction of sex offenders supervised, treated and polygraph tested in the Oregon and Washington correctional treatment systems over the same time period, and so these results may not be generalizable. The sample cases were delimited by the availability of data for every adult male that was evaluated, treated, and polygraph tested at program offices during those years, as a result of the principal author´s role as the former outpatient program director.

Descriptive Statistics

The sample cases consisted of 170 men, the majority of whom were Caucasian (97.6%). The mean age of offenders, at the time

of polygraph testing was 36.6 years, the median age is 34.5 years, standard deviation of 14.56 years and range of 18 to 80 years. Table 1 shows the types of sexual offenses.

Table 1 Frequency of types of sexual offenses

Sex Offender Type				
Child Molester	Rapist	Non-contact	Other	Total
n = 110	*n* = 33	*n* = 20	*n* = 7	*N* = 170

The sample cases consisted of low risk offenders (*n* = 51, 30.0%); moderate-low risk offenders (*n* = 72, 42.4%); moderate-high risk offenders (*n* = 37, 21.8%) and high risk offenders (*n* = 10, 5.9%) as classified by the Static-99R (Hanson and Thornton, 2000; Hanson, 2005). Table 2 shows the frequencies, percentages and Static-99R risk levels.

Table 2 Frequency distribution of Static-99R risk scores and risk levels

Static-99R Score		Frequency	Percent	Risk Level
-3		*n* = 3	1.8	Low
-2		*n* = 3	1.8	Low
-1		*n* = 7	4.1	Low
0		*n* = 14	8.2	Low
1		*n* = 24	14.1	Low
2		*n* = 36	21.2	Moderate-Low
3		*n* = 35	20.6	Moderate-Low
4		*n* = 26	15.3	Moderate-High
5		*n* = 12	7.1	Moderate-High
6		*n* = 6	3.5	High
7		*n* = 2	1.2	High
8		*n* = 1	0.6	High
9		*n* = 1	0.6	High
Total	100.0%			

The mean Static-99R score was 2.41, the median 2.0, standard deviation 2.057, with a range of -3 to +9. Not surprisingly, the majority of offenders were classified in the moderate risk categories. Relatively few of the sample cases were in the low and high risk categories.

Other independent control variables were also analyzed, including the presence of substantial denial, sexual deviance and psychopathy. Denial of the instant offense details and/or problematic sexual behavior at intake was observed and documented by the evaluator or therapist in 57.1% (*n* = 97)

of the cases. Sexual deviancy, as indicated by self-reported paraphilias during presentence investigation or psycho-sexual evaluation reports, or through penile plethysmograph testing, was a factor for 48.8% (n = 83) of the individual offenders. Antisocial personality disorder or psychopathy was diagnosed by a psychologist or other licensed mental health professional and described in the presentence or psycho-sexual evaluation reports for 12.9% (n = 22) of the offenders.

Sample variables related to the this study were also analyzed, including treatment outcome, disclosure of sexual offense history, and whether a non-deceptive sexual history polygraph was completed within six months of treatment onset. Non-deceptive sexual history polygraph results were achieved by 67.1%, (*n* = 114) of the 170 sample cases. For 47.1% (*n* = 80) cases the non-deceptive sexual history polygraph was completed within six months of treatment onset. Case records indicated that 32.9% (*n* = 56) did not achieve a non-deceptive sexual history polygraph examination result. Offenders who completed treatment comprised 41.8% (*n* = 71) of the sample. A small portion of the sample 2.9% (*n* = 5) received no treatment due to the recommendations contained in the presentence and psychosexual evaluation reports. Some of the offenders did not complete treatment, including those who were terminated for non-compliance 17.1% (*n* = 29), dropouts 5.3% (*n* = 9) and discharges due to jurisdiction or funding changes 26.5% (*n* = 45). Some cases 5.3% (*n* = 9) were transferred to other treatment programs, and 1.2% (*n* = 2) deceased. Some cases were counted in multiple categories.

Findings

Table 3 shows the recidivism offenses committed by 39 recidivists, including failure to register/report, sexual abuse, child molestation, and forcible rape. Eleven (6.5% of the total sample) recidivists perpetrated new sex crimes within 5 years of discharge from treatment program discharge. Twenty eight (22.9% of the total sample) recidivists failed to register or report, a status sexual offenses. Within the group of status reoffenders, 50% (n = 14) also committed new non-sexual crimes, including non-sexual assault (n = 6), felony possession of controlled substance (n = 3), and theft of varying degrees (n = 5).

Table 3 Frequency of recidivism during a 5-year follow-up period

Sexual Recidivism Offense	Frequency
Child Molestation offense	*n* = 1
Dealing Child Pornography	*n* = 1
Encouraging Child Sexual Abuse	*n* = 1
Failure to Register/Report (FTR)	*n* = 28
Incest offense - Sexual Exploitation of Minor	*n* = 1
Rape	*n* = 2
Child Sexual Abuse/Assault	*n* = 5

Total 5-Year Recidivists (including FTR)	*n* = 39 (22.9% of *N* = 170)
5-Year Sexual Recidivists (excluding FTR)	*n* = 11 (6.5% of *N* = 170)

Results of bivariate analyses are shown in Table 5. That analysis showed statistically significant relationships between sexual recidivism and the variables of non-deceptive sexual history polygraph result (r_ϕ = .167, *p* = .029) and age under 35 at time of non-deceptive sexual history polygraph result (r_ϕ = -.152, *p* = .047). That analysis also showed a statistically significant relationship between successful completion of sex offense specific treatment and a non-deceptive sexual history polygraph result at any time during treatment (r_ϕ = .328, *p* < .001). There was also a statistically significant negative correlation between sexual deviancy

and treatment completion (r_{ϕ} = -.168, p = .028). Other study variables, including denial, were not statistically significantly correlated with treatment completion or recidivism, as indicated in Table 4. Only the Static-99R variable was approaching a significant level (r_{ϕ} = .134, p = .082) relative to sexual recidivism.

Table 4 Correlation of study variables to treatment completion and sexual recidivism

Variable	Treatment Completion Value / Sig. Level	Sexual Recidivism Value / Sig level
Age under 35 at non-deceptive SHPE	-.095 / .217	.167* / .029
Denial	-.057 / .461	.035 / .649
Non-deceptive SHPE within 6 months	.077 / .317	-.152* / .047
Non-deceptive SHPE	.328* / .000	-.070 / .361
Psychopathy	.041 / .592	-.101 / .186
Sexual Recidivism	.004 / .959	----------
Sexual Deviance	-.168* / .028	.126 / .101
Static 99R Risk Score	-.129 / .095	.134 / .082
Treatment Completion	----------	.004 / .959
Total Cases N = 170		

*Significant at p < .05

Non-deceptive SHPE results were correlated with successful completion of treatment (r_{ϕ} = .382, p < .001), but the correlation with sexual recidivism was not statistically significant. There was a statistically significant negative relationship between a non-deceptive SHPE within six months of treatment onset and sexual recidivism (r_{ϕ} = -.152, p = .047) at the .05 level. Interestingly, age under 35 at time of non-deceptive SHPE was also significant for sexual recidivism at the .05 level (r_{ϕ} = .167, p = .029). Sexual deviancy was negatively correlated with completion of treatment (r_{ϕ} = -.168, p = .028). No adjustment or correction was made to theses reported p-values.

Table 5 shows the case frequencies for the study variables, and indicates that nine of 11 sexual recidivist cases (81.8%) did not complete a non-deceptive sexual history polygraph within six months of treatment onset, and this relationship was statistically significant (r_{ϕ} = -.152, p = .047). Results also show that 81% of sexual recidivists were under age 35 at the of the SHPE, and that the relationship between age under 35 at the time of a non-deceptive sexual history polygraph examination and sexual recidivism was statistically significant (r_{ϕ} = .167, p = .029). Cases from the two moderate risk groups, as determined by Static-99R risk scores, were most represented in the recidivist category (63.6%), though the relationship risk and recidivism was not statistically significant (r_{pb} = .134, p = .082) for this sample.

Table 5 Frequency of sexual recidivism by study variables

		Non-deceptive SHPE within 6 months of treatment onset		
		No	Yes	Total
Sexual Recidivism	No	81	78	159
	Yes	**9**	**2**	**11**
		Age at time of non-deceptive SHPE		
		<35	>35	
Sexual Recidivism	No	83	76	159
	Yes	**9**	**2**	**11**

		Static 99-R Risk Level				
		Low	Med Low	Med High	High	Total
Sexual Recidivism	No	49	68	34	8	159
	Yes	**2**	**4**	**3**	**2**	**11**

Table 6 shows the crosstab frequencies for each independent variable and whether a non-deceptive SHPE was completed within six months of treatment onset. Forty-eight (60%) of the 80 offenders who completed a non-deceptive SHPE within six months of treatment onset were under age 35, which is statistically significant (r_{ϕ} = 0.189, p = .014). Denial was a factor for 69 of 90 cases for which a non-deceptive SHPE was not completed within six months of treatment onset, while denial was a factor for only 35 of 80 (35%) of the offenders who completed a non-deceptive SHPE within six months of treatment onset (r_{ϕ} = - .420, p = .000). No statistical correction was used for these reported p-values. Correlations were not statistically significant for the relationship between non-deceptive SHPE results and five other study variables, including recommendations for continued treatment at the time of discharge, sexual deviancy, anti-social personality or psychopathy, and Static-99R risk level.

Table 6 Distribution for non-deceptive sexual history polygraph within six months of treatment onset

	Age at time of non-deceptive SHPE				
			<35	>35	Total
Non-deceptive SHPE within 6 months of treatment onset					
	No		37	53	90
	Yes		**48***	32	80
	Treatment Status at Discharge – More Tx Needed				
			Yes	No	Total
Non-deceptive SHPE within 6 months of treatment onset					
	No		50	40	90
	Yes		41	39	80
	Presence of Sexual Deviance				
			Yes	No	Total
Non-deceptive SHPE within 6 months of treatment onset					
	No		42	48	90
	Yes		41	39	80
	Presence of Psychopathy – APD				
			Yes	No	Total
Non-deceptive SHPE within 6 months of treatment onset					
	No		14	76	90
	Yes		8	72	80
	Presence of Denial at Intake				
			Yes	No	Total
Non-deceptive SHPE within 6 months of treatment onset					
	No		69	21	90
	Yes		**28***	52	80
	Static 99-R Risk Level				
		Low	M-Low	M-High	High	Total
Non-deceptive SHPE within 6 months of treatment onset						
	No	31	32	22	5	90
	Yes	20	40	15	5	80

*Significant at $p < .05$

Discussion

This project was a correlation study, involving N = 170 adult males who were convicted of sexual offenses, designed to investigate the relationships between treatment and recidivism outcomes and other variables, including the completion of a non-deceptive sexual history polygraph examination, under age 35 at time of non-deceptive sexual history polygraph examination, anti-social personality disorder or psychopathy, sexual deviancy, and denial of offense details at the time of treatment intake. Bivariate analysis suggests that some study variables are worthy of further interest and continued investigation of their value towards the prediction of sexual recidivism and treatment completion. Two variables were significantly correlated with treatment completion, non-deceptive sexual history polygraph examination results (r_ϕ = .328, p < .001) and sexual deviancy (r_ϕ = -.168, p = .028), though the relationships of these and the sexual recidivism outcome were not statistically significant. Other sexual history polygraph variables were significantly correlated with sexual recidivism: completion of non-deceptive polygraph examination results within six months of treatment onset (r_ϕ = -.152, p = .049) and age under 35 at the time of a non-deceptive sexual history polygraph exam (r_ϕ = .167, p = .029). Although non-deceptive sexual history polygraph result was associated with lower sexual recidivism, non-deceptive sexual history polygraph results by offenders under age 35 were associated with higher sexual recidivism.

In a very preliminary way, these results indicate a statistical relationship between sexual history polygraph examination results and sexual recidivism outcomes. These results suggest the possibility of an interaction between age and motivation for disclosure and the outcome of sexual recidivism might be an interesting area for further study. However, data were not analyzed for younger age adults who completed a non-deceptive sexual history polygraph within six months of treatment onset. Although the non-random sample is somewhat small and the stability or generalizability of these results is presently unknown, this bivariate analysis suggests the possibility that treatment outcomes and 5-year sexual recidivism rates may be partially informed by timeliness of non-deceptive sexual history polygraph test results. Further study is needed to better understand the value or meaning of non-deceptive sexual history polygraph results when evaluating motivation and treatment progress.

These results must be interpreted cautiously because sexual history polygraph is a complex process with a number of dimensions. Non-deceptive sexual history polygraph results indicate that all an offender produced statistically significant truthful numerical scores for all investigation target questions. Test results are neither deterministic nor a direct physical measurement of the amorphous social construct of deception. Instead, test results are a probabilistic computation of the margin of error or level of confidence that can reasonably assigned to a categorical conclusion of deception or truth-telling when comparing the numerical scores of validated physiological discriminators with statistical reference distributions for deceptive and truthful persons.

Target questions for sexual history polygraph exams are selected for their actuarial, operational, or clinical relevance to risk assessment, risk management and treatment goals. Target issues for sexual history polygraph testing are presently unstandardized, though the basic requirements for polygraph questions are that relevant target questions describe a behavioral issue that can be answered either *yes* or *no,* and for which the examinee will know whether the answer is truthful or deceptive. Relevant questions will often include this following: sexual offenses against underage children since becoming an adult, sexual contact with relatives and family members, forced sexual contact/violent sexual offenses, and sexual contact with persons who were asleep or unconscious. Other target questions can be used to investigate sexual behaviors that can signal problems involving sexual compulsivity or sexual preoccupation, including: voyeurism, exhibitionism, public masturbation, stalking behaviors, theft or use of underwear/undergarments or personal property for masturbation or sexual arousal, child pornography, and any other sexual behaviors that may indicate problems with sexual deviancy.

Achievement of a non-deceptive sexual history polygraph result necessitates that the offender first admit the allegation of the instant offense, as there are potential error hazards associated with attempting to screen for unknown sexual assault behaviors while denying a alleged sexual offense. Without strong evidence that denial of the instant offense is factually truthful, any attempt to

screen for unreported sexual offenses might amount to a form of collusion with the offender. Preparation for sexual history polygraph testing will involve reviewing conceptual vocabulary terms that describe sexual abuse behaviors, along with operational definitions that define and describe and those behaviors. Preparation will also involve a personal review of one's history of sexual behavior with the goal of identifying behaviors that were abusive or unlawful and those that were within normal limits. Preparation of and review for sexual history testing is a clinical process for which the details will be inform and be informed by the larger clinical treatment picture, including corresponding mental health and personal trauma issues, in addition to the nature and extent of the individual's history of sexual behaviors.

Effective preparation for sexual history polygraph testing may be partially a function of the quality of professional and therapeutic rapport between the offender and the treatment and supervision professionals. This preparation may also be a function of the treatment cohort group, family system, and social support network. Ultimately, effective preparation will contribute to structured and organized review of the sexual history during the sexual history polygraph pretest interview.

An important aspect of the sexual history polygraph is that offenders are not compelled to disclose identifying information regarding their sexual assaults. Instead they are permitted to withhold information about jurisdiction, exact name or the exact nature of the relationship. This is viewed as unfortunate by some, but is neither intended to devalue the impact of sexual on the personhood of sexual abuse victims, nor to endorse the rights of the offender as more important, but is necessary to facilitate the disclosure of more complete information without compelling such disclosure that it results in legal vulnerability related individual rights.

Sexual history polygraph testing is a multidimensional process that may be affected by a number of factors. Sexual history polygraph testing, as a component of sex offender treatment and management, especially in cognitive behavioral programs that encourage the reduction of cognitive distortions and the increase of personal responsibility, may be an indicator of motivation for learning and change. Polygraph testing may be used by some programs as tool for assessing readiness for activities and privileges such as social/ public events, safety and activity plans, and family reunification. Finally, the achievement of non-deceptive sexual history polygraph results may also be a function of program expectations, for which both overt and subtle message can either reinforce or undermine attitudes and perceptions about the meaning and value of the disclosure process.

In this small study, age under 35 at the time of a non-deceptive sexual history polygraph was found to be correlated with the outcome of sexual recidivism at a minimal statistically significant level. As indicated in Table 6, 60% of the offenders who completed a non-deceptive sexual history polygraph within six months of treatment onset were under age 35, 35% of those were in denial at intake, and 75% were in the two lowest Static-99R risk categories. These results suggest the potential that some difference may exist for the under age 35 group, which also appears to be over-represented in this sample. Unfortunately, for this study, we did not capture information about the details of the offenders' reported histories of sexual offenses.

Limitations

The most significant limitations to this study relate to the type of sample, sample size and project design. This study was conducted on a convenience sample of 170 adult male offenders referred to outpatient treatment programs in the Pacific Northwest region of the United States between 1994 and 2004--and who were administered at least one sexual history polygraph examination. This study was designed to be an investigatory, correlational survey aimed at evaluating whether the variable of non-deceptive sexual history polygraph results are associated with sexual recidivism. This research was not an experimental study and so it did not include a non-polygraphed control group. Causal inferences are not possible based on these results, and the generalizability of these findings is presently unknown.

Interaction effects were not evaluated in this correlation study, as might have been accomplished with data coded in a manner that would support a multivariate analysis. Multivariate analysis would have also relieved concerns about the effects of multiplicity in a survey study of significant relationships between numbers of variables. Another, previously mentioned, limitation to the design of this project has to do with the absence of

data on the nature and scope of the reported sexual offense behaviors, both prior to and during the polygraph testing process.

Recommendations

Future studies should include this type of data into the design and analysis, perhaps using level of detail similar to that outlined in Pratley and Goodman-Delahunty (2011), which includes incidents of abuse, duration of offending (in days), frequency of offending, number of locations, range of abusive acts committed, and intrusiveness of abuse. Caution should always be exercised in the area of professional expectations for disclosure of information, always recognizing the impossibility of the notion of *full disclosure* or the expectation that professionals can somehow know *everything* or *every detail* of an offender's history of sexually abusive behavior. Instead, it remains within the realm of realistic possibility that non-deceptive sexual history polygraph results signify only the probability that an offender has reported the major behavioral detail as described by the test stimuli, and that there may always remain additional under-reported detail regarding behavioral and interactional aspects of sexual abuse behaviors.

For future research, the evaluation of a larger and more randomly selected sample will be important. Comparison of recidivism outcomes for a polygraph cohort and non-polygraph cohort may also be informative of the value of polygraph testing. Future studies should attempt to capture more fine-grained data and information about the length of time necessary to achieve a non-deceptive sexual history polygraph test result, and the contribution of CRDs as a recidivism predication variable in regression or other form of prediction model. Other research questions might include qualitative aspects of how supervision and treatment professionals know, or measure, the extent to which their client is being honest in CRDs, and what additional information can be accessed by supervision and treatment professionals to more effectively predict treatment and sexual recidivism outcomes.

Replication of this study is recommended, including the comparison of these results with polygraph cohorts and non-polygraph cohorts for whom reported sexual history is verified in some other manner. Future studies should attempt to further investigate the possibility that causal relationships may exist between motivation for disclosure and outcomes for both treatment and sexual recidivism. It remains possible that the observed results are an anomaly resulting from programatic or sampling factors. At present, the potential value for sexual history polygraph testing appears worthy of a recommendation for continued interest and continued study.

References

Abrams, S. (1991). The use of polygraphy with sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 4*(4), 239-263. doi:10.1177/107906329100400401

Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 12*(2), 123-138.

American Polygraph Association. (2007). Model policy for post-conviction sex offender testing. *Polygraph, 36*(3).

Association for the Treatment of Sexual Abusers (ATSA). (2004). *Practice standards and guidelines for the evaluation, treatment and management of adult male sexual abusers.* Retrieved April 8, 2009, from https://www.atsa.com/pubSoT.html

Ben-Shakhar, G. (2008). The case against the use of polygraph examinations to monitor post-conviction sex offenders. *Legal & Criminological Psychology, 13*(2), 191-207. doi:10.1348/135532508X298577.

Bourke, M. L. & Hernandez, A. E. (2009). The 'Butner Study' Redux: A report of the incidence of hands-on child victimization by child pornography offenders. Journal of Family Violence, 24, 183-191.

Cook, R. (2011). Predicting recidivism of the convicted sexual offender using the polygraph and Static-99. (Doctoral Dissertation). Retrieved from UMI Dissertation Publishing, ProQuest, (UMI Number: 3445252).

Emerick, R.L. & Dutton, W.A. (1993). The effect of polygraphy on the self-report of adolescent sex offenders: Implications for risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, 6, 83-103.

English, K. (1998). The containment approach: An aggressive strategy for the community management of adult sex offenders. *Psychology, Public Policy and Law, 4*(1/2), 218-235.

English, K., Jones, L., Pasini-Hill, D., Patrick, D., & Cooley-Towell, S. (2000). *The value of polygraph testing in sex offender management. research report submitted to the National Institute of Justice*. Denver, CO: Denver Colorado Department of Public Safety, Division of Criminal Justice, Office of Research and Statistics.

Farber, B.A. (2003). Patient self-disclosure: A review of the research. *Journal of Clinical Psychology*, 59 (5), 589-600.

Farber, B.A. & Hall, D. (2002). Disclosure to therapists: What is and is not discussed in psychotherapy. *Journal of Clinical Psychology*, 58 (4), p 359-370.

Furedy, J. J. (1996). The North American polygraph and psychophysiology: Disinterested, uninterested, and interested perspectives. *International Journal of Psychophysiology, 21*(2-3), 97-105. doi:DOI: 10.1016/0167-8760(96)00003-7

Gannon, T.A., Wood, J.L., Pina, A., Tyler, N., Barnoux, M.F.L., & Vasquez, E.A. (2014). An evaluation of mandatory polygraph testing for sexual offenders in the United Kingdom. *Sexual Abuse: A Journal of Research and Treatment, 26 (2) 178-203.*

Grubin, D. (2008). The case for polygraph testing sex offenders. *Legal & Criminological Psychology, 13*, 177-189. doi:10.1348/135532508X295165

Grubin, D., Madsen, L., Parsons, S., Sosnowski, D., & Warberg, B. (2004). *Prospective study of the impact of polygraphy on high-risk behaviors in adult sex offenders.* United States: Retrieved from National Criminal Justice Reference Service Abstracts.

Handler, M., Nelson, R., & Blalock, B. (2008). A focused polygraph technique for PCSOT and law enforcement screening programs. *Polygraph*, 37(2), 100-111.

Harris, A., Phenix, A., Hanson, R. K. & Thornton, D. (2003). Static-99 Coding Rules: Revised 2003, the Static-99R. Ottawa, ON: Solicitor General of Canada. Printed by the Justice Institute of British Columbia, Corrections and Community Justice Division.

Heil, P., Ahlmeyer, S. & Simons, D. (2003). Crossover sexual offenses. *Sexual Abuse: A Journal of Research and Treatment, 15*(4).

Helmus, L., Babchishin, K. M., Hanson, R. K. & Thornton, D. (2009). Static-99R: Revised age weights. Retrieved January 10, 2011 from: www.static99.org

Hindman, J., & Peters, J. (2001). Polygraph testing leads to better understanding adult and juvenile sex offenders. *Federal Probation, 65*(3), N. PAG.

Honts, C. R. (2004). The psychophysiological detection of deception. In P. A. Granhag, & L. A. Stromwall (Eds.), *The detection of deception in forensic contexts* (pp. 103-123). Cambridge, England: Cambridge University Press.

Honts, C. R. (1994). Psychophysiological detection of deception. *Current Directions in Psychological Science (Wiley-Blackwell), 3*(3), 77-82. doi:10.1111/1467-8721.ep10770427.

Honts, C. R. (1996). Criterion development and validity of the CQT in field application. *Journal of General Psychology, 123*(4), 309.

Honts, C. R., & Alloway, W. R. (2007). Information does not affect the validity of a comparison question test. *Legal & Criminological Psychology, 12*(2), 311-320. doi:10.1348/135532506X123770

Honts, C. R., Hodes, R. L., & Raskin, D. C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70*(1), 177-187. doi:DOI: 10.1037/0021-9010.70.1.177

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology, 79*(2), 252-259.

Iacono, W. G. (2008). Accuracy of polygraph techniques: Problems using confessions to determine ground truth. *Physiology & Behavior, 95*(1-2), 24-26. doi: DOI: 10.1016/j.physbeh.2008.06.001

Iacono, W. G., & Lykken, D. T. (1997). The validity of the lie detector: Two surveys of scientific opinion. *Journal of Applied Psychology, 82*(3), 426-433.

Justice Institute of British Columbia – JIBC. (2011). Static-99R: Sex Offender Risk Assessment. Retrieved January 10, 2011 from http://www.jibc.ca/course/soap105

Kokish, R., Levenson, J.S., & Blasingame, G. D. (2005). Post-conviction sex offender polygraph examination: Client-reported perceptions of utility and accuracy. *Sexual Abuse: A Journal of Research and Treatment, 17*(2).

Krapohl, D. (2006). Validated polygraph techniques. *Polygraph, 35*(6), 149.

Levenson, J.S. (2009). Sex offender polygraph examination: An evidence-based case management tool for social workers. *Journal of Evidence-based Social Work*, 6: 361-375.

LexisNexis Accurint (2011). Public record criminal history records search program description. Retrieved May 9, 2011 from: http://www.accurint.com

Marshall, W.L., Thornton, D., Marshall, L.E., Fernandez, Y., & Mann, R. (2001). Treatment of sexual offenders who are in categorical denial: A pilot project. *Sexual Abuse: A Journal of Research & Treatment*, 13, 205-215.

McGrath, R.J., Cumming, G.F., Burchard, B.L., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.

McGrath, R. J., Cumming, G., Hoke, S. E., & Bonn-Miller. (2007). Outcomes in a community sex offender treatment program: A comparison between polygraphed and matched non-polygraphed offenders. *Sexual Abuse: A Journal of Research and Treatment, 19*(4), 381.

National Research Council of the National Academies of Science, Committee to Review the Scientific Evidence on Polygraph (Ed.). (2003). *The polygraph and lie detection* (D.O. Sciences ed.). Washington DC: The National Academies Press.

Nelson, R, Handler, M., Krapohl, D., Gougler, M., Shaw, P. & Bierman, L. – Ad-hoc Committee on Validated Techniques – American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. *Polygraph* 40 (4).

O'Connell, M. A. (1998). Using polygraph testing to assess deviant sexual history of sex offenders. (Doctoral Dissertation), University of Washington, Seattle, WA. Retrieved from Dissertation Abstracts International (49).

O'Connell, M.A. (2011). Recommendations for research methodology to avoid conflict of interest, dual relationships and vested interest: The utilization of independent third party research assistants. Personal communication on February 26, 2011.

Offe, H., & Offe, S. (2007). The comparison question test: Does it work and if so how? *Law and Human Behavior, 31*(3), 291.

Oregon Judicial Department. (2009). Oregon judicial information network [OJIN], access to Oregon cases on the internet. Retrieved October 3, 2010, from http://courts.oregon.gov/OJD/OnlineServices/OJIN/index.page

Patrick, C.J. & Iacono, W.G. (1989). Psychopathy, threat and polygraph test accuracy. *Journal of Applied Psychology*, 74, 347-355.

Pratley, J. & Goodman-Delahunty, J. (2011). Increased self-disclosure of offending by intrafamilial child sex offenders. *Sexual Abuse in Australia and New Zealand*, 3 (1) 10-22.

Rosky, J.W. (2013). The (f)utility of post-conviction polygraph testing. Sexual Abuse: A Journal of Research and Treatment, 25(3), 259-281.

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.

Tabachnick, B. C., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.

Washington State Institute for Public Policy, Director Roxanne Leib. (2009). *Recidivism study technical assistance – obtaining criminal history data.* Retrieved December 1, 2009 from http://www.wsipp.wa.gov/default.asp

Washington State Patrol (2011), Washington Access to Criminal History – WATCH, Retrieved May 1, 2011 from: https://fortress.wa.gov/wsp/watch.

Wilcox, D. T., Sosnowski, D., & Middleton, D. (1999). The use of the polygraph in the community supervision of sex offenders. *Probation Journal, 46*, 234-240.

Wilcox, D. T., & Sosnowski, D. E. (2005). Polygraph examination of British sexual offenders: A pilot study on sexual history disclosure testing. *Journal of Sexual Aggression, 11*(1), 3-25. doi:10.1080/13552600410001667797

A new paradigm for the experimental study of Malintent

Charles R. Honts

Psychology Department

Boise State University

Abstract

A new laboratory paradigm for the study of credibility assessment and deception concerning malintent was tested. Malintent may be a distinct concept from the traditional deception for past action that has been the subject of numerous studies. Sixty participants were either innocent or were given information and malintent to commit a theft. Participants were then screened for malintent using a modified Test for Espionage and Sabotage (TES). The TES was scored with the Kircher and Raskin (1988) discriminant analysis algorithm and produced better than chance results that were comparable to the results for the TES in a forensic deception detection setting. (Honts & Alloway, 2007). This new paradigm provides an experimental framework for exploring the concept of malintent and efforts to detect it.

Key Words: *malintent, deception detection, national security screening, portals*

A large body of research indicates that unassisted individuals, including trained forensic and security professionals are only slightly better than chance at detection deception (Hartwig & Bond, 2011). In the post 9-11 environment, the United States Government responded to the problem of assessing credibility at portals in several ways involving new research and the application of new techniques to the field setting (Honts & Hartwig, 2014). Unfortunately, all of those efforts are scientifically problematic because of poor methodology, and because they fail to address the conceptual differences between assessing credibility for intent to perform bad acts and assessing credibility for the commission of past acts (Honts & Hartwig). Polygraph examiners who do employment screening examination in law enforcement and national security face a similar credibility assessment problem as that faced at portals. Individuals may present themselves for polygraph screening with the malintent to perform bad actions if they are hired, but they may not, at the time of the polygraph screening have actually committed any bad acts. It is not clear how polygraph screening tests would perform under such circumstances.

Honts and Hartwig (2014) note that the deceptive context for assessing malintent differs in critical ways from assessing credibility concerning a past act. Most of the research conducted on credibility assessment has focused on the problem of detecting deception concerning statements about acts that took place in the past while the study of deception for intent has received little attention (Granhag, 2010). Most research is typified by asking questions about some event that the person either did or did not participate in the the past. A typical study from the past event deception literature would have a mock crime where some participants stole money and some

Corresponding Author:

Charles R. Honts
Psychology Department
Boise State University
1910 University Drive MS-1715
Boise ID 83716-5664 USA
chonts@boisestate.edu
Voice: 208.867.2027
FAX: 208.426.4386

did not (see Honts & Reavy, 2015 for a recent example). In such a setting the guilty person has the episodic memories associated with the criminal act and a concern for deception detection. Concern about deception detection then results in a variety of sequelae that involve masking the deception, monitoring the receiver and associated physiological and behavioral responses (Vrij & Gannis, 2014). For the innocent person, the concern that her or his truthful statements will not be believed is energized by the same potential consequences and sequelae as those faced by the guilty.

The deceptive context for assessing malintent at a portal or at an employment screening session is quite different (Honts & Hartwig, 2014). At the portal, or in employment screening situation, the innocent person is not the accused in a criminal investigation, and it seems doubtful that many innocent people approaching border or transportation portals, or employment screening, feel anything near the equivalent of the emotional response felt by a falsely accused criminal suspect. However, Honts and Hartwig also note that the truthful person at a portal may well feel anxious about the general process of screening. Nevertheless, it seems likely that for most innocent individuals, approaching a portal is a necessary inconvenience that may cause the innocent aggravation and minor anxiety, but little concern of true jeopardy (Honts & Hartwig). In an employment screening situation, a person without malintent seems unlikely to have much concern about questions concerning malintent. Certainly the concern and jeopardy for truthful individuals in an employment screening situation is much less than that of a falsely accused suspect in a criminal investigation.

For those intending to do bad acts if granted access, the deceptive context may also be different from the traditional situation (Honts & Hartwig, 2014). Those intending to do bad acts, may not have, as yet, have committed any bad acts, nor may they know specifically what their future bad acts might be. Those with malintent want to pass the portal, or obtain the job, so that he or she can do bad acts in the future. The person with malintent may or may not have false credentials, but if it is their intention to do bad acts in the future, that is the central nature of their deception and the focus of a relevant credibility assessment. To date, little research has addressed credibility assessment for malintent in the deceptive context presented by portals or employment screening. It is simply not clear whether or not research done in a criminal/investigative context will generalize to a malintent situation.

There is a body of research concerning deception detection for intent. Typical of this research was a study by Vrij, Leal, Mann, and Granhag (2011). In Vrij, et al., participants were asked to pretend they were part of a mission to collect a package from a specified location and then deliver it somewhere else. Participants were told that they might be stopped by agents and were given a code exchange that would identify an agent as friendly or hostile. Participant were instructed to be truthful with friendly agents and to lie about their intent and mission to the hostile agents. While on the way to complete the mission participants were stopped and interviewed by either a friendly or a hostile agent. After completing the mission participants were stopped a second time and interviewed by a friendly or hostile agent. Data extracted from the interviews indicated that after the completed event were more markers of plausibility and they contained more details than the interviews about intention. However, when undergraduate college students were asked to evaluate transcripts of the interviews no significant effects of intention versus completed event were found.

Although Vrij et al., tested the detection of intent, I would argue that they did not test malintent because in both the truthful and deceptive conditions every participant had exactly the same knowledge. While this created a powerful experiment design, it is not representative of conditions in the field. The lack of any specific motivation associated with deception detection is also troubling. Kircher, Horowitz and Raskin (1988) reported a meta-analysis that found explicit motivation was an important variable in predicting laboratory accuracy rates with polygraph tests. Some other studies that tested intent, but without malintent or specific motivation associated with deception detection were reported by Meijer, Verschuere, and Merckelbach (2010), Sooniste, Granhag, Knieps and Vrij (2013), and Vrij, Granhag, Mann, and Leal, (2011).

The paradigm described in this study was an attempt to model the critical aspects of malintent detection in a laboratory paradigm. For purely pragmatic reasons we decided to use psychophysiological deception detection (PDD) methods and technology. There is a substantial literature on the use of PDD to assess credibility in both forensic and screening settings (Raskin & Kircher, 2014). Although clearly not perfect, PDD has consistently shown better

than chance performance in both forensic and screening settings and provides a substantial amount of information gain over unassisted deception across a wide range of base rates (Honts & Schweinle, 2009). Although PDD in its present form is clearly not applicable to airport portals, it is used in national security, and law enforcement employment screening situations. Moreover, the substantial database of deception detection in traditional settings provides a stable reference for initial efforts to detect malintent.

Method

Participants

Participants were 60 college students enrolled in General Psychology classes. The participants received course credit for their participation and as part of the manipulation described below some participants had the possibility to win $14 in movie passes. Participants average age was 20.7 years, *SD* = 3.16, and 37 (62%) of the participants were men.

Apparatus

The apparatus was the same as that used in Honts and Alloway (2007). Physiological data were collected with a Stoelting commercial polygraph instrument running version 3.2 of the Computerized Polygraph System (CPS) software (Kircher & Raskin, 2002). Respiration data were collected from Pneumotrace sensors placed over the upper chest and the abdomen. Skin conductance was collected from two Ag-AgCl electrodes placed on the palmar surface of the distal phalanx of the first and third finger of the participant's left hand. Relative blood pressure was recorded from an inflated cuff placed around the participant's upper right arm. Vasomotor activity was recorded from the palmar surface of the participant's left thumb. Movement was monitored with a sensor placed under the legs of the participant's chair. The CPS software was used to edit artifacts from data and evaluate the data.

Procedure

Participants were recruited through the General Psychology course signup software. They were responding to an ad for a research project where participants might be asked to lie and then take a polygraph test. On arrival, participants watched a video about their rights as research participants and were then given an informed consent form to review and sign. The video and form indicated to the participants that they could withdraw from participation at any time and still received their participation credit in General Psychology.

After agreeing to continue, the research assistant instructed the participants chose an envelope from a box. The research assistant then left the participant alone in the room. Participants thus opened that envelope in private. The envelope contained instructions to watch a second video and a password for that video. Two participation videos defined the two conditions of the study. The Innocent Participant video provided the following instructions: Some people approaching the portal polygraph examination were given the combination to the safe in Room A and those people have the intention of opening the safe and stealing something valuable within. However, as an innocent person you do not know the combination to the safe and you have no intention to steal anything. You will be given a polygraph test about knowing the combination to the safe, and about your intention to open the safe in Room A. Maintain your innocence and lack of knowledge during the polygraph and if you pass your polygraph you will be admitted to Room A. There you will be immediately debriefed and your participation will be over. However, if you fail the polygraph you will be taken to Room B where you will watch a 30-minute educational film (TV Eye, 1983) and then you will be debriefed.

The Malintent Participant video (with the same person as shown in the Innocent Video) provided the following instructions: Some of the participants are going to be given the combination to the safe in Room A. You are one of those persons. The combination to the safe is 1 0 6 A. You will be given a polygraph about your knowledge of the combination and about your intention to rob the safe. If you are successful in passing the polygraph, you will be admitted to Room A where you can open the safe and take movie passes worth $14. You will be allowed to keep those movie passes. To pass the polygraph you must maintain your lack of knowledge and intention through the entire examination. If you fail the polygraph you will be taken to Room B where you will watch a 30-minute educational video and then you will be debriefed. Failing the polygraph means you will never have the chance to take the movie passes.

Using methods similar to those used

by Honts and Alloway (2007) all Participants were then tested with a standard field polygraph instrument and they were given a complete polygraph using a modified version of Test for Espionage and Sabotage (TES; Department of Defense Polygraph Institute Research Division Staff, 1999). The TES was chosen so that the results of our malintent polygraph could be directly compared to the results Honts and Alloway. The relevant questions of this examination were: Do you know the combination to the safe in Room A? Do you have a plan to open the Safe in Room A? If you are admitted to Room A, will you attempt to open the safe? Are you taking part in a plan to take something from the safe in Room A? Physiological data were collected from three repetitions of each relevant question. Following data collection the examiner used the discriminant analysis classification algorithm included in the CPS software (Kircher and Raskin, 1988; 2002) to classify the participant as truthful or deceptive. Depending upon the classification the participant was then taken to the room associated with their outcome. Participants who passed their examinations were given the movie passes and those who produced deceptive results on their examinations were asked to watch the educational film. Participants were then debriefed by a research assistant and were fully informed about the design of the experiment. Participant questions were answered.

Results

The discriminant analysis procedure used in this study (Kircher & Raskin, 1988); provided as part of the CPS software) produced an a posteriori probability of truthfulness (p|T) that can be considered by itself for its direct informative value or used against a cut score for classification. Those $p|T$ values were then tested to see if deception detection was possible in this setting. An independent groups t-test of the $p|T$ values indicated that significant detection was obtained, $t(58) = 3.26$, $p = 0.002$. The $p|T$ for Innocent participants ($M = 0.69$, $SD = 0.37$) was higher than for Malintent participants ($M = 0.39$, $SD = 0.34$). The correlation between the guilt criterion and the $p|T$ values was 0.394, $p < .01$. These results are similar to those reported by Honts and Alloway (2007) for the TES in a forensic setting (Innocent $M = 0.72$ and Guilty $M = 0.40$). If decisions were made so that $p|T > .5$ were classified as truthful and $p|T$ values $< .5$ were deceptive, 70% of the malintent and 67% of the innocent individuals were classified correctly. That classification was significantly above chance, $\chi^2 = 8.01$ (1), $p = .004$, Kendall's *tau-b* = .37, $p = .002$, and was again similar to the performance of the TES in Honts and Alloway's (2007) forensic paradigm.

Discussion

The present results provide a proof of concept for a new paradigm to assess malintent. Malintent participants approached a screening task not having committed a transgression in the past, but with knowledge and intent to commit a transgression if they were able to pass the screening. Innocent participants approached the screening task without malintent or malintent related information and were motivated to pass the screening to avoid delay. A modified version of the U. S. Government's Test for Espionage and Sabotage, a psychophysiological deception detection test, was used as the credibility assessment tool. The TES was able to discriminate malintent from innocent participants at better than chance levels. The performance of the TES in this malintent paradigm was much better than that reported for unassisted individuals (Hartwig, Granhag, & Luke, 2014) and was comparable to the performance of the TES in a forensic laboratory paradigm (Honts & Alloway, 2007). Unfortunately, the nature of the equipment and the time necessary for administration of the TES make it an unreasonable candidate for use a high volume portals, although it is currently used for employment screening in national security screening settings in the United States (Department of Defense Polygraph Institute Research Division Staff, 1998).

Nevertheless, these results validate this paradigm as a way to establish a basic malintent paradigm. The paradigm is easy to implement and should be easily adaptable to a variety of manipulations that would allow for the explication of the malintent construct. Research is urgently needed to define the limits and nature of the malintent concept. Although already being widely attempted in the field (Honts & Hartwig, 2014), such basic research would seem to be absolutely necessary before legitimately applying techniques to detect malintent in the field.

References

Department of Defense Polygraph Institute Research Division Staff. (1998). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. *Polygraph*, *27*, 68–73.

Granhag, P. A. (2010). On the psycho-legal study of true and false intentions: Dangerous waters and some stepping stones. *Open Criminology Journal*, *3*(2), 3743. doi:10.2174/1874917801003020037

Hartwig, M., & Bond, C.F. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin, 137*, 643–659.

Hartwig, M., Granhag, P. A., & Luke, T. (2014). Strategic use of evidence during investigative interviews: The state of the science. In, Raskin, D. C., Honts, C. R., & Kircher, J. C. *Credibility assessment: Scientific research and applications:.* (pp. 1-36). Oxford, UK: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394433-7.00001-4

Honts, C. R. & Alloway, W. (2007). Information does not affect the validity of a comparison question test. *Legal And Criminological Psychology, 12,* 311-312. (Available online in 2006)

Honts, C. R. & Hartwig, M. (2014). Credibility assessment at portals. In, Raskin, D. C., Honts, C. R., & Kircher, J. C. *Credibility assessment: Scientific research and applications:.* (pp. 37-61). Oxford, UK: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394433-7.00002-6

Honts, C. R., & Reavy, R., (2015). The comparison question polygraph test: A contrast of methods and scoring. *Physiology and Behavior, 143*, 15-26. Published online 24 February 2015, doi:10.1016/j.physbeh.2015.02.028

Honts, C. R., & Schweinle, W. (2009). Information gain of psychophysiological detection of deception in forensic and screening settings. *Applied Psychophysiology and Biofeedback, 34,* 161-172. (Available online July 2009)

Kircher, J.C., Horowitz, S.W., Raskin, D.C., 1988. Meta-analysis of mock crime studies of the control question polygraph technique. *Law and Human Behavior, 12*, 79–90

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.

Meijer, E. H., Verschuere, B., & Merckelbach, H. (2010). Detecting criminal intent with the concealed information test. *The Open Criminology Journal*, *3*, 44-47.

Raskin, D. C., & Kircher, J. C. (2014). Validity of polygraph techniques and decision methods. In, Raskin, D. C., Honts, C. R., & Kircher, J. C. *Credibility assessment: Scientific research and applications:.* (pp. 63-129). Oxford, UK: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394433-7.00003-8

Sooniste, T., Granhag, P. A., Knieps, M., & Vrij, A. (2013). True and false intentions: asking about the past to detect lies about the future. *Psychology, Crime & Law*, *19*, 673-685.

TV Eye (1983). Telling the truth. Episode in TV Eye series, London, UK: Thames Color Productions.

Vrij, A., & Gannis, G. (2014). Theories in deception and lie detection. In, Raskin, D. C., Honts, C. R., & Kircher, J. C. *Credibility assessment: Scientific research and applications:.* (pp. 301-374). Oxford, UK: Academic Press. http://dx.doi.org/10.1016/B978-0-12-394433-7.00007-5

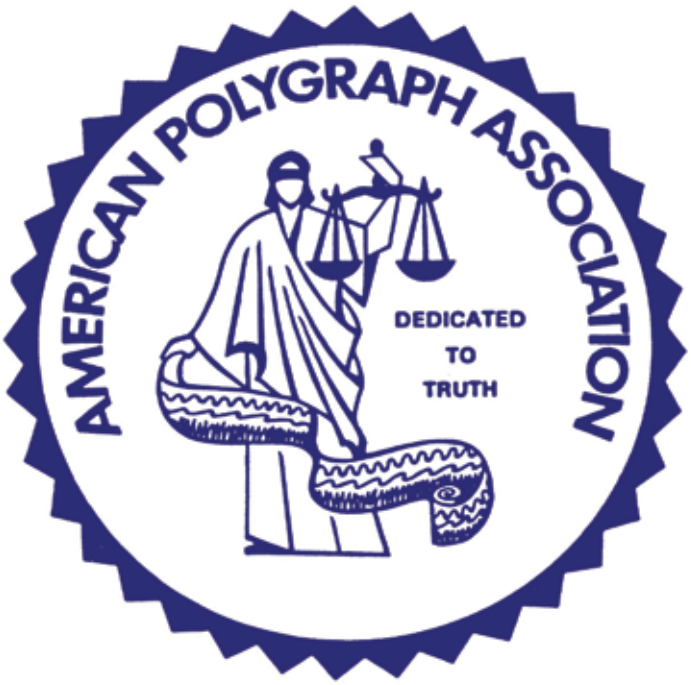
Vrij, A., Granhag, P. A., Mann, S. A., & Leal, S. (2011). Lying about flying: The first experiment to detect false intent. *Psychology, Crime & Law*, *17*, 611-620.

Vrij, A., Leal, S., Mann, S. A., & Granhag, P. A. (2011). A comparison between lying about intentions and past activities: Verbal cues and detection accuracy. *Applied Cognitive Psychology*, *25*, 212–218. http://dx.doi.org/10.1002/acp. 1665.

Author Note

The author would like to thank the following individuals for their work in collecting the data in this study: Scott McBride, Flavia Pittman, James Pittman, Adela Anderson, and Ashley Christiansen. Correspondence concerning this paper should be sent to the author at: Psychology Department, Boise State University, 1910 University Drive, Boise, ID 83725-1715 or chonts@boisestate.edu

Some of the results presented here were also given as a poster presented at the annual meeting of the Association for Psychological Science in Chicago: Honts, C. R., Pittman, F. A., Pittman, J. V., McBride, S. T., Anderson, A. B., & Christiansen, A. K., (2008, May). *A New Paradigm for the Study of Deception Detection at Portals*.


AMERICAN POLYGRAPH ASSOCIATION
DEDICATED
TO
TRUTH