EUROPEAN

POLYGRAPH

PUBLISHED QUARTERLY

Volume 10 2017 Number 4 (38)

Andrzej Frycz Modrzewski Krakow University

DE GRUYTER
OPEN

EUROPEAN
POLYGRAPH

Volume 10 • 2016 • Number 4 (38)

DOI: 10.1515/ep-2016-0021

John C. Kircher*
David C. Raskin
University of Utah,
Salt Lake City, UT, USA

Laboratory and Field Research on the Ocular-motor Deception Test

Лабораторные и практические исследования применения «ocular-motor» для выявления лжи

Key words: ocular-motor", "ocular-motor detection test, instrumental detection of deception

Disclosure: The authors have financial interests in Converus Inc. (www.converus.com), a company that has commercialized the technology described in this report. We have disclosed those interests to the University of Utah and have in place an approved plan for managing any potential conflicts that arise from involvement in Converus.

The present paper reviews the rationale and theoretical assumptions that underlie the ocular-motor deception test (ODT) as well as empirical evidence of its criterion-related validity. The research suggests that the ODT may contribute to pre-employment and periodic screening programs, particularly in government agencies concerned with law enforcement and national security.

* John.Kircher@utah.edu
** davidc.raskin@me.com

© year of first publication Author(s). This is an open access article distributed under the Creative Commons Attribution-NonCommercial-NoDerivs license http://creativecommons.Org/licenses/by-nc-nd/3.0/

Rationale underlying the Ocular-motor Deception Test (ODT)

Cook et al. (2012) introduced a new method for detecting deception called the ocular-motor deception test (ODT). In contrast to the polygraph, the ODT is automated and can be completed in approximately 40 minutes. A computer presents voice-synthesized and written instructions followed by written true/false test statements concerning the examinee's possible involvement in illicit activities. The instructions inform the examinee that if they do not answer quickly and accurately, they fail the test. The examinee then reads statements presented serially by the computer while a remote eye tracker recording eye movements and pupil size changes. The examinee presses a key on the keyboard to answer true or false. The computer processes the ocular-motor data, combines its measurements in a logistic regression equation, and classifies the individual as truthful or deceptive on the test.

The ODT uses a test format known as the Relevant Comparison Test (RCT). Originally, we developed the RCT as a new polygraph technique for use at ports of entry to screen travelers for possible trafficking of drugs and/or transporting explosives (Kircher et al., 2012). The RCT includes questions about two relevant issues (R1 and R2) that are intermixed with neutral questions, and it uses the difference between reactions to the two sets of relevant questions to determine if the examinee was truthful or deceptive to either of the relevant issues. Each relevant issue serves as a control for the other. Examinees reacting more strongly to questions concerning one of the issues are found deceptive in their answers to questions about that relevant issue. Examinees who show little or no difference in reactions to the two sets of relevant questions are considered truthful to both issues.

The ODT is based on the assumption that lying is cognitively more demanding than telling the truth. A recurrent theme in the literature on deception detection techniques (Johnson, Barnhardt, & Zhu, 2005; Kircher, 1981; Raskin, 1979; Steller, 1989; Vrij, Fisher, Mann, & Leal, 2006). In contrast to truthful people, a deceptive individual must identify questions answered truthfully and questions answered deceptively. When they recognize a question as inculpatory, they must inhibit the pre-potent truthful response and do so consistently, quickly, and accurately. While they are performing the task, deceptive individuals may also self-monitor their performance for signs revealing their deception, by either answering too slowly or making too many mistakes. The recruitment of mental resources to accomplish these additional cognitive and meta-cognitive activities could account for the observed impact on pupil dilation, eye movements, response time, and error rates (Hacker et al., 2012; Kahneman, 1973; Loewenfeld, 1999; Rayner, 1998).

The pupil reacts not only to cognitive load but also to emotional stimuli. Several investigators have reported that emotional stimuli evoke pupil responses whose magnitude depends on the intensity but not the valence of the emotional stimulus (Bradley, Micolli, Escrig, & Lang, 2008;. Hess & Polt,1960; Hess & Polt, 1964; Steinhauer, Boller, Zubin & Pearlman, 1983). Polygraph tests are based on the concept that deceptive individuals will show stronger emotional responses to test questions answered deceptively than to those answered truthfully. To the extent that emotional reactions to test questions distinguish deceptive from truthful individuals, pupil responses should reflect those differences and be diagnostic of deception. Consistent with this prediction, several investigators have reported that during concealed information and probable lie polygraph tests the pupil dilates more when people are deceptive than when they are truthful (e.g., Bradley & Janisse, 1979; Dionisio et al., 2001; Janisse & Bradley, 1980; Webb et al., 2009).

A reader who has difficulty reading or comprehending text shows more eye fixations, pupil enlargement, and longer reading times (Rayner, 1998; Rayner, Chace, Slattery & Ashby, 2006). If deceptive individuals experience greater cognitive load and difficulty processing test items than truthful individuals, we should see differences between the groups on these measures.

Mock crime laboratory research on the ODT

We have conducted a series of laboratory and field studies to determine if ocular-motor measures discriminate between truthful and deceptive individuals. The laboratory studies use a mock crime procedure that we modeled after laboratory research on polygraph techniques. Realistic mock crime experiments produce diagnostic effects on electrodermal, cardiovascular, and respiration reactions that are similar to those obtained from actual criminal suspects (Kircher, Horowitz & Raskin, 1988; Raskin & Kircher, 2014). In our experiments, we recruit participants from the university campus or the general community and randomly assign them to guilty and innocent treatment conditions. We instruct guilty participants to commit a mock crime, such as stealing an exam from a professor's office, or taking $20 from a secretary's purse, and then lie about it on the ODT. To simplify the research design, we have begun conducting experiments with one rather than two mock crimes. Participants in these experiments are led to believe that some guilty participants take an exam from a professor's office, whereas others take $20 from a secretary's purse. In actuality, all guilty participants take $20 from the purse. Because examinees in field settings usually are highly motivated to pass the test, we offer participants a substantial monetary bonus to appear innocent of the crimes.

The ODT consists of 16 True/False statements concerning the theft of the $20 (e.g. "I took the $20 from the secretary's purse."), 16 statements concerning the theft of the exam (e.g. "I did not take the exam from the professor's office."), and 16 neutral statements (e.g. "I am seated before a computer, taking a test."). In the standard format, we arrange the 48 statements so that no two statements of the same type appear in succession. The computer presents the set of 48 items 5 times with a brief break between repetitions during which the examinee answers simple arithmetic problems. Excluding the arithmetic problems, the examinee answers a total of 80 T/F statements of each type. The correct (non-incriminating) answer to half of the items of each type is True and the correct answer to the other half is False.

Ocular-motor and Behavioral Measures. The eye trackers in our experiments measured horizontal gaze position, vertical gaze position, and one or two channels of pupil diameter depending on whether the tracker recorded data from the left and right eyes or from the right eye only. The trackers stored samples of each channel at either 30Hz or 60Hz. When the tracker measured left and right eyes, we computed the mean of measures from both eyes, because the two channels were highly correlated.

Mean change in pupil size for a mock crime experiment reported by Cook et al. (2012) is shown in Figure 1. It shows the evoked pupil response for four seconds following the onset of the statement for neutral statements, statements about the theft of the cash, and statements about the exam. As predicted, guilty participants reacted more strongly to statements about the cash than the exam, whereas innocent participants showed little difference between cash and exam statements. It is noteworthy that the mean change in pupil diameter (PD) was less than 0.1 mm.

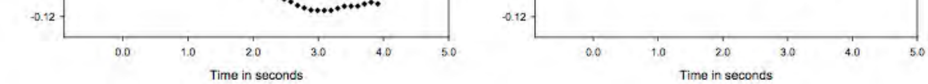

Figure 1. Mean change in pupil diameter (PD) from statement onset for guilty (left) and innocent participants (right)

To adjust for individual differences in reactivity and differences in ambient light conditions, we transformed the pupil data for each of the five repetitions of test items to standard scores. Standardization also established a common metric across repetitions within individuals. From the standardized pupil response curves, we extracted two features: the area under the response curve and the level of the response at the participant's answer.

We derived reading measures from analyses of eye fixations on the text. A computer identified eye fixations on the text where there was little variance in both horizontal and vertical gaze positions for a minimum duration of 100ms and a maximum duration of 1000ms (Cook et al., 2012). An area of interest was defined for each test item prior to the calculation of the reading measures. The area of interest began with the first character of the item and ended after the period at the end of the item. From the set of fixations for a test item, the computer measured three features:

- **number of fixations** was a simple count of the fixations in the area of interest
- **first pass duration** was the total duration of all fixations in the area of interest until a fixation fell outside the area of interest, and
- **reread duration** was the total duration of all fixations in the area of interest associated with leftward eye movements.

We divided the three reading measures by the number of characters to adjust for differences in the length of test statements.

Behavioral measures included participants' *response times* and *error rates*. We also measured blink rates per second for each item (*item blink rate*) and for the subsequent item (*next item blink rate*). Based on pioneering work by Stern, Walrath and Goldstein (1984), we expected a decrease in item blink rate when the participant was highly focused on reading an incriminating statement and an increase in blink rate when the participant encountered a less incriminating subsequent item (*next item blink rate*).

For a given type of measurement, such as pupil size at the time of the participant's answer, we calculated two contrasts. We subtracted the pupil size for statements concerning the theft of the exam (R2) from the pupil size for statements concerning the theft of the $20 (R1); i.e. (R1 – R2). The other contrast was between the mean reaction to statements answered truthfully (neutral and R2 statements) and statements answered deceptively by guilty participants (R1); i.e. (R1 – (neutral + R2) / 2). If the rationale underlying the ODT is correct, we should see relatively high or low scores on each of these contrasts for guilty participants and scores near zero for innocent participants.

We assessed the ability of ocular-motor measures to discriminate between guilty and innocent participants by correlating within-subject contrasts between statement types with guilt status, where guilt status was a dichotomous variable that was either 0 if the participant was innocent or 1 if the participant was guilty.

In addition to measuring the diagnostic validity of various ocular-motor measures, we also use coefficient alpha to assess the reliability of those measures (Chronbach, 1951). Reliability indicates the extent to which the measurements obtained from the five repetitions of the test items are consistent. For example, if the data from the first repetition indicated that the individual was deceptive, did the person also appear deceptive in the second, third, fourth, and fifth repetitions?

Accuracy of the ODT in laboratory studies

Table 1 presents validity and internal consistency (reliability) coefficients for the set of features in experiments conducted by two of our students (Patnaik, 2015; Webb, 2008). It also shows internal consistency reliability statistics for the computer-generated features.

Table 1. Validity and reliability coefficients for two mock crime studies of the ODT

			Webb (2008)		Patnaik (2015)		Mean	
	Area under the curve	R1 - R2	**0.409**	0.640	**0.586**	0.615	**0.505**	0.628
	Area under the curve	R1 - (NT+R2)/2	**0.396**	0.759	**0.554**	0.639	**0.482**	0.699
	Level at answer	R1 - R2	**0.557**	0.465	**0.585**	0.510	**0.571**	0.488
	Level at answer	R1 - (NT+R2)/2	**0.548**	0.527	**0.634**	0.575	**0.593**	0.551
Reading								
	Number of fixations	R1 - R2	**-0.509**	0.572	**-0.406**	0.627	**-0.460**	0.600
	Number of fixations	R1 - (NT+R2)/2	**-0.329**	0.807	**-0.293**	0.720	**-0.312**	0.764
	First pass duration	R1 - R2	**-0.549**	0.582	**-0.253**	0.540	**-0.427**	0.561
	First pass duration	R1 - (NT+R2)/2	-0.293	0.622	-0.166	0.585	**-0.238**	0.604
	Reread duration	R1 - R2	**-0.488**	0.516	**-0.342**	0.397	**-0.421**	0.457
	Reread duration	R1 - (NT+R2)/2	**-0.224**	0.683	-0.115	0.407	**-0.178**	0.545
Behavioral								
	Response time	R1 - R2	**-0.529**	0.434	**-0.497**	0.329	**-0.513**	0.382
	Response time	R1 - (NT+R2)/2	**-0.312**	0.788	**-0.348**	0.671	**-0.330**	0.730
	Error rate	R1 - R2	0.082	0.052	0.093	0.209	**0.088**	0.131
	Error rate	R1 - (NT+R2)/2	**0.242**	0.741	-0.002	0.690	**0.171**	0.716
Blink rate								
	Item blink rate	R1 - R2	-0.014	0.247	**-0.388**	0.182	**-0.275**	0.215
	Item blink rate	R1 - (NT+R2)/2	-0.015	0.572	-0.191	0.101	-0.135	0.337
	Next item blink rate	R1 - R2	0.169	0.104	-0.088	0.351	0.135	0.228
	Next item blink rate	R1 - (NT+R2)/2	0.010	0.315	-0.105	0.381	0.075	0.348

Bolded validity coefficients are statistically significant at $p < 0.05$.

The strong positive correlations for pupil features indicated that guilty participants showed greater increases in pupil size in response to R1 (cash) than to R2 (exam) statements. These findings are consistent with the data presented in Figure 1. The negative correlations for reading and response time measures indicated that guilty participants made fewer fixations and spent less time reading R1 than R2 statements. The effects on response time are substantial and consistent over multiple experiments and cultural groups. Based on the psychology of reading literature, we initially thought that guilty participants would experience more difficulty and spend more time on R1 than R2 statements. However, the data suggest that guilty participants invest more mental effort in answering those statements quickly and accurately in an attempt to avoid detection (Cook et al., 2012). Guilty participants achieve their objective but reveal their deception. This hypothesis explains the effects on response time and reading measures as well as the observed increases in pupil size associated with R1 statements.

Examination of mean validity coefficients indicates that the pupil measures were more diagnostic than reading, behavioral, and blink rate measures. The (R1 – R2) contrast for response time was almost as diagnostic as pupil size. Blink rate and error rate measures were the least predictive of guilt status.

Decision Model. To classify individuals as truthful or deceptive, we combine the scores on a subset of diagnostic measures in a mathematically optimal manner to compute the probability of deception. If the probability of deception exceeds 0.5, we classify the person as deceptive; if the probability is less than 0.5, we classify the person as truthful. Several statistical procedures have been developed to identify a subset of diagnostic measures that will represent most of the diagnostic variance in the full set of measures and work well when tested on an independent sample of cases. These procedures tend to select measures that are more highly correlated with guilt status and less highly correlated with each other. We then use logistic regression analysis to derive a unique weight for each ocular-motor measure that maximizes the separation between truthful and deceptive groups.

Much of our research has been designed to assess the effects of factors that could affect the accuracy of the ODT and to explore alternative methods for presenting test items. In regard to the latter objective, we have not improved on the presentation format and mock crime procedures evaluated in our first mock crime experiment (Osher, 2005), which we call the standard protocol. Because our attempts to improve on the standard protocol have yielded inferior results, Table 2 presents the results obtained with the standard protocol, and Table 3 presents results for non-standard protocols.

Table 2. Percent of correct decisions under standard conditions in mock crime experiments

Experiment	Factors	N	n_G	n_I	Guilty	Innocent	Mean	Validation $_G$	Validation $_I$	Mean
Osher (2005)[a]	issues; serial format	40	20	20	85.0	85.0	85.0	85.0	70.0	77.5
Webb (2008)[b]	sex; motivation; difficulty	112	56	56	82.1	89.2	85.7	89.3	80.4	84.9
Patnaik (2013)[a]	direct interrogation	48	24	24	83.3	95.8	89.6	83.3	83.3	83.3
Patnaik (2015)[a]	distributed; pretest feedback; post-response interval	80	40	40	82.5	90.0	86.3	80.0	90.0	85.0
Patnaik et al. (2016)[c]	language; culture	145	82	63	84.1	87.3	85.7	81.9	87.5	84.7
Middle East (2016a)	language; culture	112	51	61	80.4	88.5	84.5			
Middle East (2016b)[d]	language; culture	101	52	49				75.0	85.7	80.4
Standard Protocol		638	325	313	82.8	89.0	85.9	82.1	84.1	83.1

[a] Validation results were obtained with the leave-one-out procedure

[b] The decision model based on Patnaik et al. (2016) was used to classify participants in Webb (2008)

[c] The decision model based on Webb (2008) was used to classify participants in Patnaik et al. (2016)

[d] The decision model based on Middle East (2016a) was used to classify participants in Middle East (2016b)

The results presented in Table 2 indicate that the standard protocol in mock crime experiments yielded approximately 86% correct classifications in the original, standardization sample, and approximately 83% correct when tested on independent samples (cross-validation). On cross-validation, accuracy was slightly higher for innocent (84.1%) than guilty participants (82.1%).

Table 3 summarizes results from non-standard conditions. The results from Osher (2005) suggest that we obtain more diagnostic information from serial presentations of individual test statements than with the simultaneous display of multiple test statements. Webb (2008) found that the person's sex does not moderate the effects of deception on ocular-motor measures, whereas higher motivation to pass the test and semantic simplicity in the phrasing of test statements improves the diagnostic validity of some ocular-motor measures.

Table 3. Percent of correct decisions under non-standard conditions in mock crime experiments

Experiment	Factors	N	n_G	n_I	Guilty	Innocent	Mean
Osher (2005)	issues; parallel format	40	20	20	70.0	95.0	82.5
USTAR (2010)	pretest questionnaire; issues	71	47	27	59.6	77.8	68.7
NSA (2012)	standardization	94	51	43	72.5	88.4	80.5
NSA (2013)	validation	60	34	26	50.0	80.8	65.4
Patnaik (2013)	indirect interrogation	48	24	24	58.3	79.2	68.8
Patnaik (2015)	blocked	80	40	40	77.5	85.0	81.3
Non-standard protocols		393	216	180	65.3	84.5	74.9

Together, the USTAR (2010) and Patnaik (2013) studies indicated that test statements that refer directly to the matter at hand ("I did not take the $20.") produce higher accuracies than statements that ask indirectly if the person falsified information to cover up their guilt ("I did not falsify my answers to questions about the theft of the $20."). In the NSA studies, we recruited employees and tested them about minor security violations. The studies used a non-standard protocol because we were not permitted to provide incentives for government employees to pass the ODT, many participants were federal polygraph examiners who knew that there were essentially no consequences to failing the ODT, and we had to rely on self-report as a proxy for ground truth.

Patnaik (2015) found that the standard pseudo-random sequencing of NT, R1, and R2 statements improves the diagnostic validity of ocular-motor measures, whereas feedback about performance on a pre-ODT practice test and lengthening the interval between the answer and the presentation of the next item does not. Patnaik et al. (2016) found that the effects on ocular-motor measures were similar for tests administered in English or Spanish to native speakers enrolled as university students. The experiments in the Middle East required modification of the display software to present Arabic text from right to left. Accuracy rates on cross-validation in the Middle East were slightly lower than those obtained in the USA and Mexico, particularly for guilty participants. Although the differences in decision accuracy between Middle Eastern and Western participants were not statistically significant, we found it necessary to reduce the number of repetitions of test statements for measures of pupil response to achieve comparable levels of discrimination between truthful and deceptive Arabic-speaking participants as we had found for English- and Spanish-speaking participants. We are uncertain why it was necessary to make those changes.

Field study of the ODT

We have recently completed a field validity study of the ODT that evaluated applicants for positions in the office of Mexico Attorney General's, immigration, and federal police. We tested some applicants for recent use of illegal drugs (R1) compared to either corruption, arms trafficking, or affiliation with a religious terrorist organization (R2). Ground truth for deceptive cases were admissions by applicants during a subsequent polygraph test that they had used illegal drugs in the period covered by the statements on the ODT (n = 71). Lack of an admission is insufficient to establish conclusively that an individual was truthful on the ODT. To obtain data from truthful people, we created a new test for applicants for positions in immigration that asked if they had committed espionage (R1) or sabotage (R2). We assumed that all of the tested individuals were truthful in their answers to both relevant issues because the base rate of deception on those issues in Mexico is very low, especially for people who had no prior government employment (n = 83).

To develop and validate a decision model with the field data, we extracted ocular-motor measures from the eye tracker data. We computed validity coefficients for the measures and used linear regression to select a subset of four measures to distinguish between the confirmed truthful and deceptive groups. We then used the selected variables in a 5-fold validation of a binary logistic regression model to classify cases as truthful or deceptive. To conduct the 5-fold validation, we divided the sample of 154 field cases into five random subsamples such that each subsample consisted of approximately 20% of the deceptive cases (n=14 or 15) and 20% of the truthful cases (n=16 or 17). The first subsample of 14 truthful cases and 16 deceptive cases (N=30) was removed, and a decision model was created with the remaining four subsamples (N=124). We used that decision model to classify the holdout sample of 30 cases and recorded the percent correct for each group. The second subsample was then set aside (N=30), a new decision model was developed with the remaining 124 cases, and the accuracy of classifications was calculated for the second holdout sample. We repeated this process for the remaining three subsamples. The results are brought together in Table 4.

Table 4. Percent of correct decisions in five decision models on independent subsamples

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Mean
	N=30	N=30	N=31	N=31	N=32		N=154
truthful	75.0	87.5	88.2	88.2	100.0	87.8	86.1
deceptive	100.0	71.4	85.7	78.6	86.7	84.5	

Percent of correct decisions varied between the five subsamples from 75% to 100% for truthful applicants and from 71.4% to 100% for deceptive applicants. Our best estimates of the performance of the model based on all 154 cases when tested on a new sample of field cases are the means for truthful (87.8%) and deceptive (84.5%) applicants. Although the results obtained in the 5-fold validation using actual applicants for positions in the Mexican government suggest that the ODT may add value to a pre-employment screening program, to some degree our methods may have capitalized on chance and produced accuracy rates that are too optimistic. Specifically, we used the entire sample to select the subset of ocular-motor measures for the decision model that was subsequently validated. This particular subset of measures worked well in the 5-fold validation but might be suboptimal for a new set of confirmed field cases. For this reason, we recommend that the current decision model be re-evaluated with new cases from representative field settings.

Conclusions

The results of laboratory and field research indicate that the ODT yields accuracy greater than 80% on both truthful and deceptive examinees, although the accuracy rates tend to be 3% to 6% higher for truthful than for deceptive individuals. The accuracy rates generalize across English and Spanish languages, US and Mexican cultures, and to a lesser extent, Arabic in the Middle East. We believe the ODT to be a promising new technology that is best suited for screening applications. We also believe it is better suited to screening applications than specific-incident, criminal investigation, because it would be difficult to construct a RCT that contains two non-overlapping relevant issues with face validity. For the same reason, polygraph examiners rarely if ever use the Guilt Complex question for specific-incident polygraph examinations (Office of Technology Assessment, 1983).

Unpublished pilot research with poor readers indicated, as expected, that the ODT was ineffective. It appears that the cognitive difficulties experienced by examinees who struggle to comprehend test statements overshadow effects of deception on ocular-motor measures. We are exploring audio-visual and audio-only alternatives to the current text-based ODT. However, as yet we have not developed a functional audio version of the test that might be effective for a broader population of individuals, including those with poor reading skills.

With specialized training and practice, polygraph tests can be defeated (Honts, 2012). To date, there have been no attempts to investigate effects of countermeasures on the ODT. However, because examinees are under pressure to respond quickly and accurately, attempts to implement countermeasures may be ineffective or easily detected with behavioral measures, such as response time and error rates. Additional research is needed to explore these possibilities.

Finally, research on the ODT has primarily been conducted in a single laboratory by one team of investigators. As high quality eye tracking equipment and data analysis software become available at lower costs, we are hopeful that other investigators will contribute new knowledge to this area of applied science.

References

Bradley, M.M., Micolli, L., Escrig, M.A., Lang, P.J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45,* pp. 602–607.

Bradley, M. T., & Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: Cardiovascular, electrodermal, and pupillary measures. *Psychophysiology, 18,* pp. 307–315.

Cook, A.E., Hacker, D.J., Webb, AK., Osher, D., Kristjansson, S., Woltz, D.J., & Kircher, J.C. (2012). Lyin' Eyes: Ocular-motor Measures of Reading Reveal Deception. *Journal of Experimental Psychology: Applied, 18(3),* pp. 301–313.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16, pp. 297–334.

Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology, 38,* pp. 205–211.

Hacker, D.J., Kuhlman, B., & Kircher, J.C., Cook, A.E., & Woltz, D.J. (2014). Detecting deception using ocular metrics during reading. In D.C. Raskin, C.R. Honts, & J.C. Kircher (Eds.), *Credibility assessment: Scientific research and applications.* Elsevier, pp 159–216.

Hess, E.H., Polt, J.M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132,* pp. 349–350.

Hess, E.H., Polt, J.M. (1964). Pupil size in relation to mental activity during simple problem solving. *Science, 143,* –1190–1192.

Honts, C.R. (2012). Countermeasures and credibility assessment. In In DC Raskin, CR Honts, & JC Kircher (Eds.), *Credibility assessment: Scientific research and applications.* Elsevier, pp. 131–156.

Johnson, R., Jr., Barnhardt, J., & Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: An event-related brain potential study. *Cognitive Brain Research, 24*, pp. 386–404.

Kahneman, D. (1973). *Attention and effort.* Prentice-Hall: New Jersey.

Kircher, J. C. (1981). *Psychophysiological processes in the detection of deception.* Department of Psychology, University of Utah, Salt Lake City, UT. Unpublished manuscript.

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*, pp. 79–90.

Kircher, J.C., Raskin, D.C., Honts, C.R., & Horowitz, S.W. (1994). Genereralizability of statistical classifiers for the detection of deception. *Psychophysiology, 31*, S73. (Abstract)

Loewenfeld, I. E. (1999). *The pupil: Anatomy, physiology, and clinical applications (Vol. 1).* Boston: Butterworth-Heinemann.

Middle East (2016a). *Ocular-motor detection of deception with Middle Eastern college students, Phase 1.* Unpublished research. University of Utah, Department of Educational Psychology.

Middle East (*2016b*). *Ocular-motor detection of deception with Middle Eastern college students, Phase 2.* Unpublished research. University of Utah, Department of Educational Psychology.

NSA (2012). *Ocular-motor detection of deception with employees at the National Security Agency Phase 1.* Unpublished research. University of Utah, Department of Educational Psychology.

NSA (2013). *Ocular-motor detection of deception with employees at the National Security Agency Phase 2.* Unpublished research. University of Utah, Department of Educational Psychology.

Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation.* OTA-TM-H-15. Washington, DC: US Government Printing Office.

Osher, D. (2005). *Multimethod assessment of deception: Oculomotor movement, pupil size, and response time measures.* Unpublished dissertation, University of Utah, Department of Educational Psychology.

Patnaik, P. (2013). *Ocular-motor methods for detecting deception: Direct versus indirect interrogation.* Unpublished master's thesis, University of Utah, Department of Educational Psychology.

Patnaik, P. (2015). *Oculomotor methods for detecting deception: Effects of practice feedback and blocking.* Unpublished dissertation, University of Utah, Department of Educational Psychology.

Patnaik, P., Woltz, D.J., Hacker, D.J., Cook, A.E., Ramm, M.L., Webb, A.K., & Kircher, J.C. (2016). Generalizability of an ocular-motor test for deception to a Mexican population. International *Journal of Applied Psychology, 6(1),* pp. 1–9.

Raskin, D.C. & Kircher, J.C. (2014). Validity of polygraph techniques and decision methods. In D.C. Raskin, C.R. Honts, & J.C. Kircher (Eds.), *Credibility assessment: Scientific research and applications.* Elsevier. pp. 63–129.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* pp. 372–422.

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading, 10,* pp. 241–255.

Steinhauer S. R., Boller F., Zubin J., Pearlman S. (1983). Pupillary dilation to emotional visual stimuli revisited. *Psychophysiology, 20,* p. 472.

Steller, M. (1987). *Psychophysiologische Aussagebeurteilung* [Psychological assessment]. Hogrefe: Gottingen.

Stern J.A., Walrath L.C., Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology 21,* pp. 22–33.

USTAR (2010). Oculomotor deception detection. Unpublished research funded by the Utah Science Technology and Research Initiative.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10,* pp. 141–142.

Webb, A.K. (2008). *Effects of Motivation, and Item Difficulty on Oculomotor and Behavioral Measures* of Deception. Unpublished dissertation, University of Utah, Department of Educational Psychology.

Webb, A. K, Honts, C. R., Kircher, J. C., Bernhardt, P.C., & Cook, A. E. (2009). Effectiveness of pupil diameter in a probable-lie comparison question test for deception. *Legal and Criminal Psychology, 14(2),* pp. 279–292.