Polygraph

VOLUME 32 2003 NUMBER 1

Contents

Published Quarterly

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

Polygraph

Editor-in-Chief: Stuart M. Senter, Ph.D.

******

Associate Editors: Norman Ansley, Troy Brown, Ph.D., Andrew Dollins, Ph.D., Kim English, Frank Horvath, Ph.D., Murray Kleiner, Donald Krapohl, Vance MacLaren, Dean Pollina, Ph.D., Michael Sakuma, Ph.D., Stuart Senter, Ph.D., Shirley Sturm, Douglas Vakoch, Ph.D., Gordon L. Vaughan, Esq., Jennifer Vendemia, Ph.D., Virgil Williams, Ph.D., Lee Wurm, Ph.D.

Subscription Information: *Polygraph* is published quarterly by the American Polygraph Association. Advertising and Editorial Address is P.O. Box 10342, Ft. Jackson, SC 29207 (USA). Subscription Rates: One year $80.00 (domestic), $100.00 (foreign). Change of address: APA National Office, P.O. Box 8037, Chattanooga, TN 37414-0037.
The Sheridan Press

Editorial

Stuart M. Senter

Polygraphy is a unique profession, representing a demanding blend of talents and proficiencies, with connections to many varied areas. These include public safety, law, national security, and adjudicative decisions, to name a few. We have a unique blend of membership, including government, law enforcement, and private polygraph examiners, lawyers, and scientists. Thus, it comes as no surprise that *Polygraph* historically has contained a broad spectrum of topics and readership interest. This trend will continue, with the ultimate goal of education. To achieve this goal, I intend to include new and current items, in addition to works from the past to maintain a sense of history, and convey an understanding of where we have been as a profession, where we currently stand, and our vision of where we will go in the future. Staying informed and educated, concerning all aspects of the psychophysiological detection of deception, is the key to our continued success as a profession. As always, knowledge is power, and my vision is for *Polygraph* to serve as a vehicle to that end.

I offer the sincerest gratitude to the two prior editors of the American Polygraph Association, Dean Pollina and Don Krapohl. Both of these individuals continue to offer me their insight and wisdom, both as Associate Editors and as sounding boards for my ideas and aspirations for *Polygraph*. I consider myself fortunate to be privy to the wealth of information and experience that they offer, and I hope to continue their legacy with a rational balance of science and application.

I give thanks in advance to all of the selfless Associate Editors who volunteer their valuable time and effort in this endeavor. It is these individuals who help to ensure that you are provided with a useful and useable product. They deserve the deepest of appreciation from us all. I also want to thank all of those who have, who are, and who will send us items for consideration. Thank you for taking the time to share your knowledge and the fruits of your efforts.

I encourage the submission of any scholarly work, including research articles, legal briefs, instructional materials, bibliographies, literature reviews, book reviews, case histories, and any other conceivable items that may serve to educate and elevate our profession. Any such items will be considered for publication. Following the lead of the Editors before me, I continue to encourage electronic submission of manuscripts and other items for consideration, as this facilitates the review, editorial, and ultimately the publication process. If you require any assistance, technical or otherwise in this capacity, please do not hesitate to contact me. In addition, submissions following the writing style described in the Publication Manual of the American Psychological Association (currently on the 5th edition), where possible, are encouraged.

Finally, I consider it a great honor to serve as your Editor-in-Chief and to be a part of this great organization. I welcome any comments, suggestions, or input that you may have...and of course, your submissions! Onward and upward.

Numerical Evaluation and Wise Decisions[1]

Donald J. Krapohl, Brett A. Stern, & Yazmin Bronkema

Key words: accuracy, commentary, decision rules, features, numerical evaluation, reliability, scoring, test data analysis, theory, validity

In the first part of the 20th century, rules for the interpretation of physiological data in lie detection were appearing in scientific journals as scientists began to explore the use of bodily responses for this purpose. The level of scientific attention in the succeeding decades was uneven, but never strong. As a result, many or most of the polygraph chart interpretation rules that found their way into current polygraph practice were developed by the practitioners and sages of the profession, rather than through rigorous scientific methods. These practitioner rules now largely dominate the field, and have been repeated in various books, school handouts, seminar materials, and polygraph publications. A keyword search of the articles in the technical journal *Polygraph* over the last 30 years finds 123 articles have appeared on the topic in that one publication alone, with the majority of these reiterating beliefs of the authors rather than producing convincing data.

When one examines the larger psychophysiological literature, much seems to be known about physiological responding. There is far less agreement among polygraphers, however, as to the best way to evaluate charts. Disagreements and public debates over scoring and decision rules have persisted for decades: which numerical threshold or cutting scores give the best accuracy; against which comparison question should the relevant question be scored, is hyperventilation (or some other particular feature) diagnostic, etc? Despite a lack of consensus among polygraph professionals, these and other scoring questions actually do have correct and defensible answers. They are founded on empirical research and rational decision-making, rather than the historical approach of personal preferences and school traditions. In this paper we will review scientific principles and findings, with an eye toward identifying those that lead to the best results.

We warn that some of the conclusions in this paper will contend with the different schools of thought regarding chart interpretation. This is not necessarily by design, but rather the result of the approach to the problem we have taken. Instead of a recitation of a particular system's scoring rules, or criticism of someone's method, or the introduction of a new means of evaluation, we looked at scoring from a fresh perspective: how can scientific principles be brought to bear on the issue of polygraph accuracy? We base our approach on well-established principles that relate to diagnostic techniques of all types, from medical tests to interpretation of satellite images, from latent fingerprint analysis to, of course, polygraphy.

Our emphasis on the science should not be taken that polygraph data analysis is exclusively a scientific undertaking. In polygraphy or other fields, some small degree of art is helpful.

[1]. This article is one in a series under the heading *Best Practices*. The first and second authors are Federal examiners. The third author is Director of International Sales with the Lafayette Instrument Company. The opinions expressed in this article are not necessarily those of the US Government or Lafayette Instrument Company. Reprint requests should be sent to: Donald Krapohl, PO Box 10411, Ft. Jackson, SC 29207, or via e-mail to dkrapohl@aol.com.

However, exploiting the science before the art permits us to minimize idiosyncratic preferences, abandon unproductive rules, facilitate the training of new examiners, and help move the field toward more general acceptance. Our goal is to urge polygraphers to examine their assumptions about polygraphy, to consider what our sister sciences have to offer us, and to invite the profession to take advantage of those principles and scientific evidence that can improve our practices individually and collectively.

In that vein, let us first discuss some important concepts.

Concepts

Validity

In order to appreciate how scoring practices affect polygraph accuracy, we must understand what this term means, and how it relates to other principles. There are several forms of validity but the one most applicable to polygraph decisions is criterion-related validity. This concept refers to how well polygraph decisions match ground truth. It is generally recognized that the polygraph does not detect lies, the criterion of interest. The polygraph merely monitors and records selected physiological functioning within the human body, permitting inferences about a person's veracity. The validity of polygraphy is gauged by how well the underlying phenomenon (differential physiological arousal) predicts deception. If decisions that are based on the phenomenon afford a high degree of agreement with ground truth, it is said to have a high criterion-related validity. Conversely, if the phenomenon does not correspond with ground truth, the validity is said to be low. Consequently, the best way to have high validity is to use only phenomena that reliably occur during deception, employ a method of interpretation that identifies and weights them according to their diagnostic value, and use optimized decision rules. Inclusion of unreliable phenomena can erode accuracy, as can sub-optimal scoring and decision rules.

Reliability

A companion concept to validity is reliability. Reliability is the measure of repeatability or reproducibility. There are three forms of reliability of interest: test-retest, intra-rater, and inter-rater reliability. Test-retest reliability relates to testing the same person with the same test on multiple occasions to assess whether the same outcome is reached on all occasions. Intra-rater reliability is how often the same person agrees with himself or herself on multiple occasions with the same data. For example, are the charts interpreted the same on Monday as they were on the previous Friday? The most frequently researched form of polygraph reliability is inter-rater, that is, the degree of agreement among different evaluators who are looking at the same data. Inter-rater agreement is extremely important, and it will set the limit for how valid a technique can be. For example, if there is only 70% agreement among scorers in the field, the accuracy of the technique cannot exceed 70%. This is because at least 30% of the scorers did not get the right answer (assuming it was the 70%, and not the 30% who did get the right answer!) However, high agreement doesn't mean high accuracy. There can be high agreement and low validity. This appears to be the case with the current voice stress technologies used in deception detection. Through practice and exercises, voice stress technicians can achieve high rates of agreement, however their criterion-related validity is still poor because the underlying phenomenon is weak or not valid. In the extreme, there can be 100% agreement, and yet the evaluators can be wrong every time. It is important to recognize that high agreement is necessary for, but does not guarantee, high validity.

Variance

Most everything varies. Polygraph scores certainly do. So do IQs, body sizes, pulse rates, amount of sunshine on any given day, number of family members, reaction time, number of sweat pores on the palm, and an uncountable number of other things. Variance must be taken into account when research is done, because samples also have variance between them. The extent to which one can rely on research findings depends upon the source from which samples were taken and how large they were. We often read

articles where one scoring system had 90% accuracy while another had 80% or 100% or some other number. Sometimes authors will claim that 90% is better than 80% or worse than 100%, usually to the advantage of the author's argument. As a particularly good (or bad) example of this, in 2001 an author in *Polygraph* characterized a reduction of a mere three inconclusive decisions (from 8 to 5) as a 37.5% change, and touted this percentage as proof of the author's point of view. Had the author conducted a test of proportions, that difference would have been found to be meaningless (z=0.87, ns). This example of improper statistical methods, and others like it prompted the present paper.

Is a finding of 90% really better or worse than other percentages? Is a reduction of 37.5% something we can rely on? If the sample sizes aren't adequate, the differences are probably not very stable. That's because a different sample might have come up with entirely different percentages, possibly upending previous conclusions. If one had large samples, the finding of a difference becomes more reliable, but even then there should be a replication. Unless the research has at least been replicated with another sample, one can never be sure that the findings extend beyond the original sample. Other factors separating good science from bad science include the sampling method, criterion selection, exclusion criteria, and selection of statistical methods.

Like most everything else, the physiological responses of examinees, and the scores that result from them, also vary. Though deceptive examinees tend to have scores more in the negative direction, and non-deceptive examinees in the positive direction, there is a range and distribution of scores for these groups. Some liars produce extremely negative scores, while other liars tested on the same issue might have scores significantly more in the positive direction. The frequency of scores tends to graph into the shape of the familiar bell curve (See Figure 1). Those scores most often obtained will cluster near the middle of the distribution, and their frequency tapers off as the scores become more extreme. The most negative score obtainable in the 7-position manual scoring system with three charts and three relevant questions is -81. It is exceptionally unusual for anyone to have a score this low. Similarly, a non-deceptive person with the same number of questions and charts can receive a total of +81 points. Again, it is exceedingly rare to see scores this high. It is this relationship between scores and their frequency that produce the bell curve.

Figure 1. Bell curve.

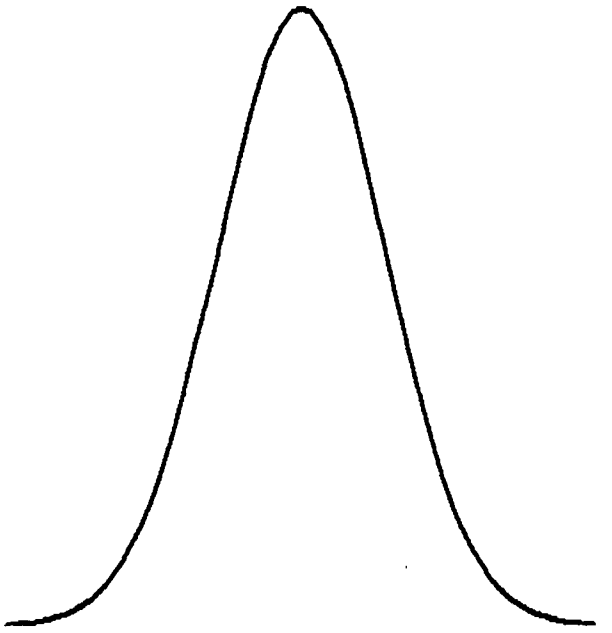


Figure 1 is the idealized bell curve, though there has been scant attention paid to whether polygraph scores produce such a perfect shape or one that is more skewed. Nevertheless, the following principles are fairly robust, and will apply to all except markedly non-normal distributions.

The evidence suggests that the scores from deceptive examinees fall mostly in one bell curve, and the scores from nondeceptive examinees in another. When a technique has poor diagnosticity, the two bell curves will be heavily overlapped. Using Figure 2 as a hypothetical example, the distribution of scores for the liars covers an area shared largely with those of truthtellers. Because of this overlap, the diagnosticity of the technique is poor, regardless of where one placed the cutting score.

Figure 2. Two hypothetical distributions, suggesting poor diagnosticity for a test.

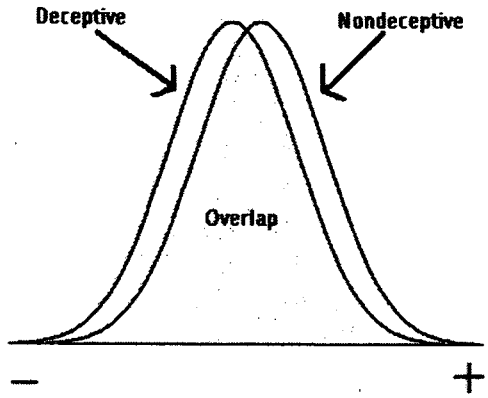
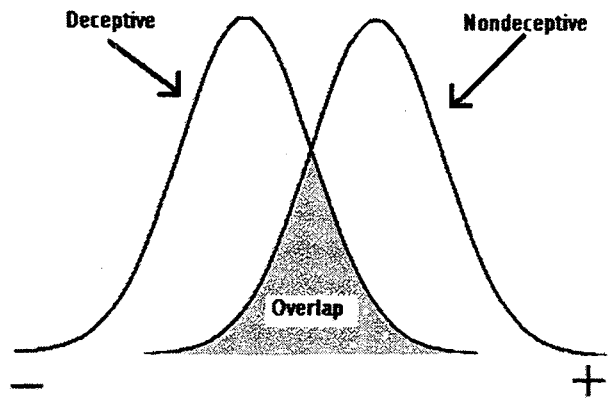


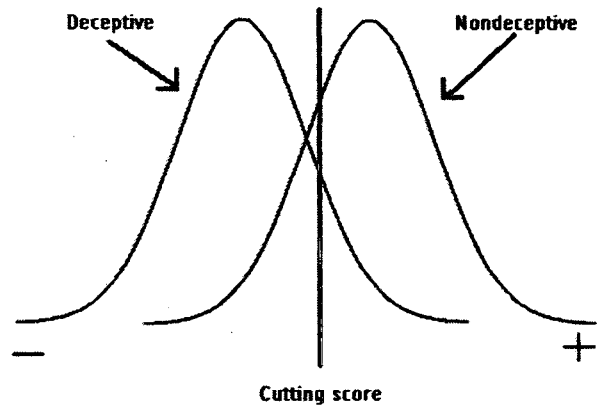
Figure 3 shows two distributions that have less overlapping area. If the method used valid features and rules to create the distributions, any cutting score used with Figure 3 would always outperform any cutting score with Figure 2 in terms of proportions of correct decisions.

Figure 3. Two hypothetical distributions indicating relatively good diagnosticity.



A careful look at these graphs will show that cutting scores do not really affect accuracy as much as they affect the type of errors that are made. For example, moving a cutting score more toward the positive direction permits the scorer to capture a larger portion of the deceptive examinees (See Figure 4). However, there is a cost of misclassifying nondeceptive examinees. Similarly, shifting the cutting score the other way helps identify more truthtellers, but causes a loss of detecting liars. Catching more of one type means losing some of the other type. Most examiners grasp this relationship intuitively. Nevertheless, there are a number of research articles from examiners reporting efforts to find "perfect" cutting scores that result in the least number of errors. As Figure 4 reveals, such efforts have been, and will always be unsuccessful because the underlying distributions always overlap in deception testing.

Figure 4. Graph depicting how adjusting a cutting score to increase detection of deception affects the detection of truthfulness.



Though shifting the cutting scores may be a poor method of improving accuracy, there are other methods that are effective. One such method is to include only the most diagnostic information in the scoring system. As more predictive physiological features are added to the polygraph scoring system, the distributions of scores of the truthful and deceptive examinees are pulled in opposite directions. As they shift away from one another, there is less overlap, and consequently less error across cutting scores. Compare Figures 2 and 3. Figure 3 shows distributions further apart, indicating that the scores were based on more predictive features than those in Figure 2, thereby improving accuracy in a way that adjusted cutting scores cannot.

A note of caution is warranted here. Merely adding more scoring features, scoring

rules, or increasingly complex processes to the mix will not move the two bell curves apart. They may even make the process of manual chart interpretation more difficult and less objective. Only valid features have a positive effect. When the number of features is much larger than perhaps a dozen or so, marginally effective or even ineffective features are creeping into the model. This is especially the case when working with physiological data, which tends to have high inter-subject variability.

If "more" meant "better" for polygraph features, one could hypothetically develop a scoring system with a thousand features, and claim that the larger number makes that system superior to those with fewer than a thousand features. The fallacy of this line of reasoning should be obvious, and it is the rarest of diagnostic fields that can claim to have more than 20 individual features that are found reliable enough for use by human evaluators. As a case in point, in a technical report to the U.S. government, Harris, Horner, and McQuarrie (2001) found that 22 of the manual scoring criteria reported by Swinford (1999) could be reduced to four and deliver the same information, the remaining features being either redundant or ineffective. Similarly, the simple three-feature Objective Scoring System enjoyed better accuracy than human scorers of the same data sets that had 22 features in their scoring system (Krapohl & McManus, 1999; Krapohl & Norris, 2000).

There is also a human factor to consider. Unlike computer algorithms, which can accommodate extremely complex calculations with perfect reliability, the reliability of human decision-making correlates with simplicity. Increasing complexity erodes reliability among human scorers, and as discussed earlier, this reliability is essential for validity. Adding rules, features, or decision rules can, at some point, diminish accuracy. It is an example of when more is less.

Next, let us consider inconclusive outcomes. How one comes to inconclusive decisions is another factor that can affect accuracy. The wider the inconclusive band, the fewer errors are made. Looking again at Figure 3, one can see that it is possible to have a scoring system that produces no errors, though the inconclusive rate might be as high as 70% or more. Having a very narrow inconclusive zone will increase the number of correct decisions, and also incorrect decisions because, numerically, more definitive decisions are being rendered.

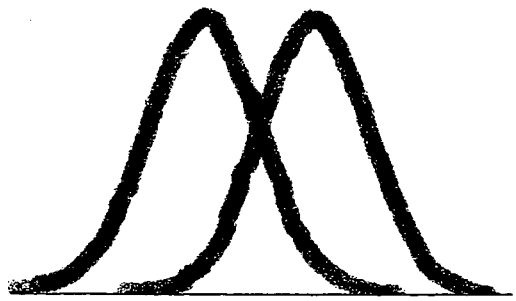
This relationship between inconclusives and error reveals that the question "How accurate is the polygraph test?" is overly simplistic. The most accurate answer to this question is, it depends. It can be as high as 100% accurate, or as low as perhaps 80%. Wide inconclusive zones decrease error, potentially approaching zero when the inconclusive area is very large. Or, with no inconclusives, accuracy is in the area of about 80%. Compare polygraphy to the field of latent fingerprints. The common wisdom is that fingerprinting produces virtually perfect decisions. The rarely discussed other side of this coin is that the technique also only produces decisions in a small minority of cases. That is because latent prints are usually of insufficient quality to be helpful. This is one reason law enforcement will only make an effort to search for them in the most serious of crimes. If polygraphy were permitted an inconclusive rate similar to that of latent fingerprinting, accuracies might well be similar.

With this understanding of the relationship among cutting scores, accuracy, and inconclusives, we are now ready to contemplate the best approach to determining numerical thresholds, or cutting scores. There are three core issues for deciding where cutting scores should be placed. They are the proportion of false positive errors, false negative errors, and inconclusive results the consumer of the polygraph results can tolerate. *Examiners can control only two of these three factors with their decision rules.* It is not possible to simultaneously have no errors and no inconclusives when a test has imperfect validity. One can choose to have low false negatives and inconclusives, but it must be paid for in false positives. Or, it is possible to have low inconclusives, but an increase in errors is inevitable. Or, a zero error rate is achievable, but the reader will know by now that the inconclusive rate will be unacceptably high.

Complicating matters more, variability also occurs among scorers. For those who have not seen it themselves, a scoring exercise at the next polygraph association meeting could be an eye-opening experience: scores can vary in some cases to produce opposite outcomes. Variability in scores is not a trivial problem. If every scorer arrived at the exact same score, at least reliability would have been achieved, and possibly validity. Perfect reliability is now only possible with systems that rely exclusively on measurements, such as with Lykken scoring (Lykken, 1959), the Objective Scoring System (Krapohl & McManus, 1999), the Rank Order Scoring System (Honts & Driscoll, 1988), and any of the automated computer algorithms. None of the semi-objective scoring systems have demonstrated the potential of achieving this reliability.

The inter-scorer variability that accompanies the semi-objective field scoring methods in common practice makes setting fixed cutting scores problematic. Scorers who come from the same training, or work in the same agency tend to have better agreement than those scorers who do not. However, even among those trained in the same methods, there are almost always differences in the composite or total scores when analyzing the same charts. This makes for a somewhat fuzzy bell curve of scores, and highlights the challenge of using universal fixed cutting scores for scorings that vary from scorer to scorer. Figure 5 characterizes the problem of cutting scores and scorer variability. As one can see, this effect reduces the reliability of any estimate of polygraph accuracy with manual scores and any set of cutting scores.

Figure 5. Two hypothetical distributions of polygraph scores showing the blurred curves produced by scorers whose scores vary for the same cases.



Levels of Rules

To maximize the efficacy of polygraph decisions, it is useful to look at the problem hierarchically. The problem begins at its base with the selection of polygraph features for scoring, followed by determining how numbers should be assigned to those features. Then, decision rules using those numbers must be formulated so that the best decisions result. It is easy to recognize that working in any other order is less effective: testing cutting scores before deciding on scoring features results in little useful information. The following sections are organized in this fashion.

Features

Scientists looking at the physiological data have found 10 polygraph tracing features reliable for manual scoring. Below are listed these individual features, along with the supporting citations from the peer reviewed literature or official government sponsored research:

Respiration

1. Suppression (including apnea)

   Barland & Raskin (1975)
   Cutrow, Parks, Lucas, & Thomas (1972)
   Harris, J.C., Horner, A. McQuarrie,D.R. (2000)
   Nakayama (1984)
   Patrick & Iaconno (1991).
   Wakamatsu & Yoshizumi (1968)

2. Increase in cycle time (decrease in the cyclic rate/slowing)

   Barland & Raskin (1975).
   Cutrow, Parks, Lucas, & Thomas (1972)
   Patrick & Iaconno (1991).

3. Change in the inhalation/exhalation ratio

   Benussi, (1914)
   Burtt, (1921a)
   Burtt, (1921b)
   Landis & Gullette (1925)

4. Baseline rise

Harris, Horner, & McQuarrie (2000)
Kircher & Raskin (1988)

Electrodermal

5. Amplitude of phasic response[2]

Harris, Horner, & McQuarrie (2000)
Kircher & Raskin (1988)
Kugelmass, et al (1968)
Patrick & Iacono (1991)
Podlesny & Truslow (1993)

6. Duration of response

Kircher & Raskin (1988)
Podlesny & Truslow (1993)

7. Complexity of response

Harris, Horner, & McQuarrie (2000)
Kircher & Raskin (1988)

Cardiovascular

8. Baseline amplitude increase

Barland & Raskin (1975)
Harris, Horner, & McQuarrie (2000)
Kircher & Raskin (1988)
Podlesny & Truslow (1993)

9. Duration of response

Harris, Horner, & McQuarrie (2000)
Kircher & Raskin (1988)

Vasomotor

10. Reduction of pulse wave amplitude

Kircher & Raskin (1988)
Patrick & Iacono (1991)

Some polygraph schools teach that there are more, sometimes dozens more, diagnostic polygraph features for manual scoring. We often hear of a polygraph examiner reporting that he or she has seen other physiological response patterns beyond these 10 on a given set of charts, and when the examiner confronted the examinee, a confession of guilt was elicited. Anecdotes and selective recollections fall far short of proof of a relationship between that particular response pattern and ground truth for most examinees, however. Though perhaps providing the makings of an interesting case, experiences like this tell little about what is generally true for all examinees. With manual numerical evaluation and current instruments, it is unlikely that more than a few important and reliable new diagnostic features will be identified by scientists any time soon.

While not amenable to human interpretation, there are data contained within the traditional polygraph channels that have shown promise as additional criteria. Foremost is respiration line length (RLL; Timm, 1982). RLL is the measure of the length of the respiration waveform over a specified period of time. RLL is a summary measure that captures respiratory suppression, change in inhalation-exhalation ratio, and increases in cycle time in a single value. So diagnostic is RLL that Harris, Horner, and McQuarrie (2000) determined that it could replace all other respiratory features currently taught. However, RLL is difficult to measure manually, and a channel that displays RLL in a meaningful way is not available on any commercially available polygraphs. For this reason, RLL is taught in only a few polygraph schools.

A second feature that has value is pulse deceleration (Patrick & Iacono, 1991). For a brief period after stimulus onset, a deceptive response is often associated with a slowing of the pulse

[2] This is but a partial list. Virtually no study has failed to find EDR amplitudes to be diagnostic.

Because most polygraphs are not currently configured to separate pulse rates from the more complex cardiograph waveform, human evaluators are less able to recognize these decelerations when they occur. With the advent of computer polygraphs, pulse deceleration and RLL could easily become additional data channels for examiners, and we are hopeful that manufacturers will see the potential in adding them.

As suggested earlier, the intractable problem that prohibits adding tracing features for manual scoring beyond the 10 listed above is that these additional features are usually suitable only for a very small number of examinees, but are irrelevant or counterindicative for most others. Take for example hyperventilation. For the rare examinee, deception may be accompanied by a noticeable increase in breathing rate for a few cycles. For virtually all other examinees, the increase in breathing rate is the type of random variation that is characteristic of physiological data in general, or perhaps a deliberate manipulation of respiration. The increased rate hardly ever means anything in terms of a person's veracity, but can be merely a normal fluctuation in breathing behavior unrelated to deception. Hyperventilation would certainly be diagnostic for some small subset of examinees, but distinguishing for whom it is diagnostic from the much larger group for which it is not is a problem no one has yet solved. Returning to the bell curves earlier in this article for clarification, because hyperventilation is not reliably diagnostic, it does not move the two bell curves apart -- it does not increase accuracy nor decrease error. It can be characterized as noise, and there are literally dozens of similar examples of individual features (e.g. premature ventricular contractions, etc.) taught in scoring systems across the profession. One day there will be additional data sensors added to the polygraph. Those sensors will be selected by how much the information they provide moves the two bell curves apart. Currently, the 10 listed above have been shown by the scientific method to be the most reliable features for manual scoring. No others taught in the field meet this high standard.

Number Assignment

There are two main approaches to manual numerical evaluation: rank order, and the 7-position numerical scale. Rank order scoring is not as widely practiced, and we will not expend much space discussing it, except to note that it does afford potentially outstanding inter-rater agreement because of its simplicity. Those interested in further reading on rank order scoring approaches are invited to read articles by Gordon and Cochetti (1987), Honts and Driscoll (1987), Krapohl, Dutton and Ryan (2001), and Miritello (1999).

Most polygraphers are familiar with the 7-position numerical scale. Traditional 7-position scoring has a notable similarity to the Likert scale (Likert, Roslow & Murphy, 1934), the tool used in psychology for over 65 years to measure attitudes. In the Likert scale, the choices are typically: strongly agree, agree, neutral, disagree, and strongly disagree. The respondent's choices are converted to scores on a 5-position scale, and attitudes thereby quantified. Analogous to the Likert scale is the polygraphic 7-position scoring: –3 (strongly deceptive), -2 (deceptive), -1 (somewhat deceptive), 0 (neutral), +1 (somewhat truthful), +2 (truthful), and +3 (strongly truthful). Though the computation rules are different for Likert and polygraph scoring scales, the longevity of the Likert number-assignment strategy is cause for reassurance that the basis of the 7-position polygraph scoring system is sound.

Though there are subtle differences in all scoring systems, a point of significant divergence among polygraph practitioners is the question of which comparison question should be used when scoring a given relevant question. The approach promoted in the Utah scoring system is to score each relevant question against the comparison question that immediately precedes it on the test. The Utah method also systematically rotates questions, to ensure each relevant question can be scored against each comparison question over the course of testing (Raskin & Honts, 2001). In the Federal method, a relevant question is scored against the stronger of two adjacent comparison questions, if the relevant question is bracketed by them (DoDPI, 2001). Otherwise, the relevant question is scored against the nearest comparison question. In the Backster system, the scoring decision relies on whether the examinee responded significantly to the relevant question (Matte,

1996). If so, it is scored against the least reactive of the adjacent comparison questions. When the relevant question does not evoke a significant physiological response, it is scored against the stronger of the two adjacent comparison questions.

These are dissimilar approaches, and one might expect that they would have different effects on the scoring data. Absent conclusive data, it would be premature to posit which of the three performs best. It may prove to be the case that none is the best, but to achieve similar accuracies for both truthful and deceptive examinees, cutting scores will be different among the methods. This possible outcome would be consistent with the overlapping bell curves premise. A research project currently underway by the present authors will attempt to find an answer.

There is at least one important difference among these approaches that is worthy of comment. For two of the methods, Backster and Federal, examiners must choose which comparison question to use based on subjective assessments of response significance and intensity. The Utah method avoids this potential source of scorer variability by using only the comparison question immediately preceding the relevant question. From a psychometric point of view, the Utah method is more scientifically defensible because it reduces individual subjective judgments. This is an attractive advantage, and one worth serious consideration.

There may also be benefit in taking the Utah question-rotation strategy an additional step, and change the positions of both the relevant and comparison questions with each chart, but in opposite directions. With each new chart, the relevant questions could be rotated forward, and the comparison questions backward, or vice versa. In addition to permitting each relevant question to be scored against a different comparison question, this method should moderate an effect on scores that may arise from within-chart habituation (Olsen & Harris, 1998). The suggested double-rotation method is more unwieldy than static question sequences, but it mitigates two possible sources of noise variance in polygraph scoring.

Some polygraph examiners score charts with the 3-position scoring system, an abbreviated version of the 7-position system. Instead of assigning scores between –3 and +3, the range of scores in the 3-position system are between –1 and +1. Because studies have shown that field polygraph examiners assign values between –1 and +1 about 80% - 90% of the time when using the 7-position method, the 3-position system is the de facto method of scoring for the majority of cases (Capps & Ansley, 1992; Krapohl, 1998). One advantage of the 3-position system is that it may reduce some of the subjectivity associated with scoring. Another is that it can be performed much faster because there are fewer judgments to make. However, because there is a more restricted range of scores with the 3-position scoring system, there is a higher proportion of inconclusive calls when the cutting scores are not adjusted. Opposite decisions between the 3- and 7-position scoring systems are exceptionally rare.

Decision Rules

Once numbers have been effectively assigned to the individual response comparisons, it is essential to use decision rules that optimize the outcomes. Readers may have noticed that we have couched our language so to leave open the possibility of something few in the profession have considered: using the same cutting scores for all circumstances will not yield maximum benefit. While examiners reading this idea have a moment to contemplate this heresy, we should like to articulate our rationale for proposing it.

Decision errors are inevitable in any diagnostic technique, including polygraphy. Use of a fixed set of cutting scores for all cases implies that the user understands and accepts the error rates these cutting scores incur, and that these cutting scores minimize the types of errors the user wants to avoid. It seems patently unlikely that the costs of errors are equal in all settings, and under certain conditions, the best decision could be to suspend judgment (i.e., make an inconclusive or No Opinion call), even when numerical scores call for a definitive decision of deception or non-deception.

There is a long-standing axiom in the polygraph profession that states, "Believe in your charts." Yet, we advocate here that even when a numerical threshold is met it might be more prudent to suspend rendering a definitive call—at least for the moment. Consider the following examples:

Example 1.

The examiner scores the test charts, and would be prepared to make a decision of No Deception Indicated (NDI) based on very positive scores. However, the examiner notes a couple of anomalies in the charts. One is that the pneumograph scores are highly positive, while the other two channels are moderately negative, though the net effect is positive composite totals when all channels are summed. Also, the respiration responses to the comparison questions appear nearly identical to one another. Should the call be NDI, or should the examiner suspend judgment?

Example 2.

The examiner tests four individuals, one of whom most certainly must have committed the crime in question. Three of the four examinees are clearly NDI by a wide margin, a call that is supported with Concealed Information Tests (CITs). The fourth examinee just barely meets the NDI numerical threshold. The CITs indicate that the fourth examinee knows more about the crime than he should. Should the call be NDI for the fourth examinee, or should the examiner suspend judgment?

Example 3.

In a murder case, the examiner very scrupulously scores the charts, which would lead to a call of Deception Indicated (DI), but by a single point. The final polygraph results will shape the prosecutor's decision as to whether to seek the death penalty. Should the call be DI, or should the examiner suspend judgment?

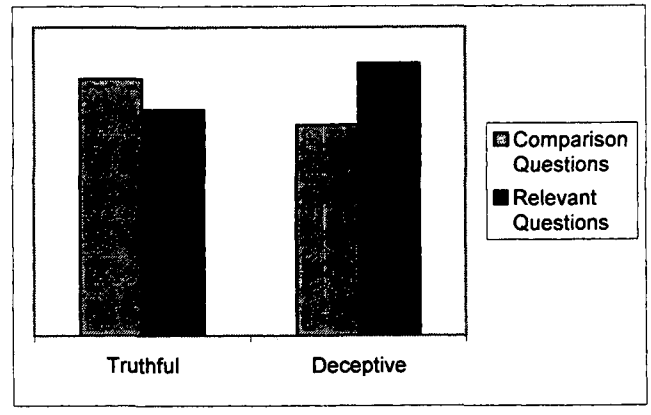
Example 4.

During the pretest interview of a pre-employment screening test the examinee reported extensive drug use. Subsequent test results warrant a call of NDI, however, the weakest analysis spot score was on the question dealing with involvement with illegal drugs. While printing the examinee's last chart, the examiner queries him for his impression of his test results, and the examinee volunteers that he believes he didn't pass the issue regarding drug involvement. Would such a disclosure warrant further exploration—perhaps through a specialized test focusing specifically on drug involvement?

Please observe from these examples that we do not suggest making opposite calls from what scoring may indicate, but that attention paid to non-scoring factors might cause the prudent examiner to withhold a decision until a retest is completed. Though more demanding of resources, waiting for the results of a retest under some circumstances could increase an examiner's accuracy.

There is another factor to consider when establishing cutting scores. Research over the last 25 years has repeatedly found that the responses of liars are not the reverse of those of truthtellers (Franz, 1989; Krapohl, 1998; Raskin, Kircher, Honts, & Horowitz, 1988). Liars, on average, give stronger reactions to relevant questions than truthtellers, on average, give to comparison questions (See Figure 6). Examiners know intuitively from experience that it is much easier to identify liars than truthtellers. Few, however, are aware that the reason is because the underlying phenomenon they are scoring, differential responding, is not symmetrical.

Figure 6. Average relative response intensity to relevant and comparison questions for truthful and deceptive examinees by individual question comparisons, in arbitrary units. Data from Krapohl, Dutton, & Ryan (2001).



Some scoring systems have cutting scores symmetrical around zero, usually +/-6. Imposing symmetrical cutting scores on an asymmetrical phenomenon means that accuracy rates will not be equal for both liars and truthtellers. One group will be detected better than the other. Because scores from field polygraph cases are normally shifted in the negative direction, symmetrical cutting scores result in better detection of liars than of truthtellers. To balance the accuracies, it would be necessary to move the cutting score for truthful decisions more toward zero, or the cutting score for DI calls more in the negative direction. However, an important lesson that bell curves teach us is that moving the cutting scores does not improve overall accuracy. It only changes the kinds of errors that result from the cutting scores.

The unbalanced accuracy rates are also affected by the Spot Score Rule (SSR). The SSR is a decision rule that depends not only on the total score of the case, but the total score of each individual question (Light, 1999). If any one question indicates deception, though it is essentially the same question as the other relevant questions, the call is DI, even if the other questions score highly positive. The SSR makes the test even more sensitive to deception, while proportionately reducing its sensitivity to truthfulness (as again predicted by the bell curves).

So, what are the best decision rules? Remember, the three factors for setting cutting scores are: the proportion of false positive errors, false negative errors, and inconclusive outcomes that the consumer can tolerate. Only two of these three can be controlled at one time. As a general rule, for fewer false negatives, the Backster system and the Federal cutting scores along with the SSR are probably the best. For more equal proportions, the Utah method is a step in the right direction. There is no scoring system for single-issue examinations in common practice with decision rules that favor the detection of truthfulness over the detection of deception. As a final thought, one might consider the available automated algorithms, some of which have been validated. Each has perfect reliability, and a validity that equals or exceeds the performance of experienced polygraphers in blind scoring polygraph charts.

Conclusion

It may come as a surprise to some that the most important player in the development of "optimal" decision rules is the consumer of the polygraph results, not polygraphers. It is the consumer who must weigh the intrinsic costs of errors and inconclusives, and it is possible to align polygraph cutting scores to correspond with the consumer's acceptance of risk. Judicious flexibility in decision rules has advantages in error containment. However, fixed decision rules have the benefit of increasing decision agreement across examiners. The fixed-threshold approach may be desirable, even necessary within agencies, though not necessarily between agencies or across the profession.

We hope we have dispelled the notion that optimization can be found simply by adding unvetted tracing features or scoring rules to the scoring system. The evidence here should make clear that byzantine manual scoring systems are to be avoided in favor of more simple and elegant methods. Diagnostic decision-making is always a difficult task, and is made more problematic by unscientific approaches sometimes seen in the field.

Polygraphy is in the proud company of medicine, psychiatry, engineering, forensic sciences, and the other fields that attempt to make critical diagnostic decisions with very complex and noisy data. With 30 years of good polygraph research to draw upon, combined with rational decision principles borrowed from other fields, the profession may be poised for the adoption of a universal scoring system, one that will be empirically tested, and will consider essential factors such as base rates, and costs and probabilities of errors. In addition, it may selectively add more automation, to increase reliability and precision. The goal is to make the process more simple, reliable, accurate, and useful than our current state of practice.

Acknowledgments

The authors wish to acknowledge the support of Mrs. Genni Arledge and Ms. Stacey Davis in obtaining reference materials and providing other assistance toward the completion of this paper.

References

Barland, G.H., & Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology, 12*, 321-330.

Benussi, V. (1914). Die atmungssymptome der lüge [The respiratory symptoms of lying]. *Archiv fuer die Gesamte Psychologie, 31*, 244-273. English translation printed in 1975 in *Polygraph, 4*(1), 52-76.

Burtt, H.E. (1921a). The inspiration/expiration ratio during truth and falsehood. *Journal of Experimental Psychology, 4*(1), 1-23.

Burtt, H.E. (1921b). Further technique for inspiration/expiration ratios. *Journal of Experimental Psychology, 4*, 106-111.

Capps, M.H., & Ansley, N. (1992). Comparison of two scoring scales. *Polygraph, 21*(1), 39-43.

Cutrow, R.J., Parks, A. Lucas, N. & Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology, 9*, 578-588.

DoDPI (2001, January). *Test Data Analysis (Manual Scoring Process): 7-Position Numerical Analysis Scale.*

Franz, M.L. (1989). *Technical report: Relative contributions of physiological recordings to detect deception.* Contract Number MDA 904-88-M-6612.

Gordon, N.J., & Cochetti, P.M. (1987). The horizontal scoring system. *Polygraph, 16*(2), 116-125.

Harris, J.C., Horner, A., McQuarrie,D.R. (2000). *An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations.* Johns Hopkins University, Applied Physics Laboratory. SSD-POR-00-7272.

Honts, C.R., & Driscoll, L.N. (1987). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph, 16*(4), 241-257.

Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*, 291-302.

Krapohl, D.J. (1998). Short report: Proposed method for scoring electrodermal responses. *Polygraph, 27*(1), 82-84

Krapohl, D.J. (1998). A comparison of 3- and 7- position scoring scales with laboratory data. *Polygraph, 27*(3), 210-218

Krapohl, D.J., Dutton, D.W., & Ryan, A.H. (2001). The Rank Order Scoring System: Replication and extension with field data. *Polygraph, 30*(3). 172-181.

Krapohl, D.J., & McManus B. (1999). An objective method for manually scoring polygraph data. *Polygraph, 28*(3), 209-222.

Kugelmass, S., Lieblich, I., Beh-Ishai, A., Opatowski, A, & Kaplan, M. (1968). Experimental evaluation of Galvanic Skin Response and blood pressure change indices during criminal interrogation. *Journal of Criminal Law, Criminology, and Police Science, 59*(4), 632-635.

Landis, C. & Gullette, R. (1925). Studies of emotional reactions: III. Systolic blood pressure and inspiration-expiration ratios. *Journal of Comparative Psychology, 5*, 221-253.

Light, G.D. (1999). Numerical evaluation of the Army zone comparison test. *Polygraph, 28*(1), 37-45.

Likert, R., Roslow, S., & Murphy, G. (1934). A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology, 5*, 228-238.

Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology, 43*(6), 385-388. Reprinted in *Polygraph, 7*(2), (1978), 123-128.

Matte, J.A. (1996). *Forensic Psychophysiology Using the Polygraph: Scientific Truth Verification – Lie Detection.* Williamsville, NY: J.A.M. Publications.

Miritello, K. (1999). Rank order analysis. *Polygraph, 28*(1), 74-76.

Nakayama, M. (1984). Suppression of respiration on the critical item and the rebound component. *Reports of the National Institute of Police Science, 40*, 32-37. Abstract in English.

Olsen, D., & Harris, J. (1998). *Study of habituation and other systematic changes in responses during field polygraph tests.* Johns Hopkins University Applied Physics Laboratory. SSD/PL-98-0368.

Patrick, C.J., & Iacono, W.G. (1991). A comparison of field and laboratory polygraphs in the detection of deception. *Psychophysiology, 28*(6), 632-638.

Podlesny, J.A., Truslow, C.M. (1993). Validity of an expanded-issue (Modified General Question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology, 78*(5). 788-797.

Raskin, D.C., & Honts, C.R. (2001). The comparison question test. In M. Kleiner (ed.) *Handbook of Polygraph Testing.* Academic Press: London.

Raskin, D.C., Kircher, J.C., Honts, C.R., & Horowitz, S.W. (1988*). A study on the validity of polygraph examinations in criminal investigations.* Final report to the National Institute of Justice. Grant No. 85-IJ-CX-0040.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven position numerical analysis scale at the DoD Polygraph Institute. *Polygraph, 28*(1), 10-27.

Timm, H.W. (1979). The effect of placebos and feedback on the detection of deception. *Dissertation Abstracts International, 40*(12-B). University Microfilm No. AAD80-13808.

Wakamatsu, T., & Yoshizumi, K. (1968). Study on respiratory wave during polygraph examination. *Reports of the National Research Institute of Police Science, 21*(3), 158-164. Abstract in English.

A Literature Review of Cross-Cultural Factors Affecting Polygraph Testing

Richard Doll, Joseph Law, and Chris Piotrowski[1]

Abstract

This report contains the results of a literary review of cultural factors which could influence the validity of polygraph examinations. Five general cultural factors were selected for study. For the purpose of this review, culture is defined by geographic area. The methodology consisted of a computer search of 11 databases. Telephone interviews were conducted with individuals actively involved in cultural research. Finally, an on-site visit to the University of Florida was conducted to allow for review of the Human Relations Area Files (HRAF). In addition to the literature review, a rational deductive approach is developed based upon Hofstede's model of cultural differentiation.

Introduction

If surveyed, the typical person on the street might provide responses such as the following: "Germans are hard working, the French are obsessed with sex, Mexicans are lazy, and Orientals are sneaky." Such descriptions are attempts at the non-scientific level to describe perceived characteristics of a culture. There is considerable research in the anthropological literature in which highly trained researchers attempt to formulate descriptions of the typical German, Frenchman, Japanese, etc. The purpose of these attempts by the layman and the scientist is to enable that person to understand the behavior and to predict the behavior of individuals whom he or she has never seen. The goal of a polygraph examiner in being able to understand the cultural effects of behavior on a potential examinee are evident. If the polygraph examiner can predict to some degree of certainty how a given individual will respond in the examination process, he or she can adjust behavior accordingly to ensure a properly administered examination and to maximize the possibility of obtaining a confession from deceptive subjects.

The problem in predicting the impact of cultural factors on behavior is complicated. Going from general predictions about a culture as a whole to specific predictions about a particular individual significantly reduces examiner's predictive powers. With these cautions in mind, one may proceed with the realization that there are common themes, beliefs, values, and attitudes which vary significantly from culture to culture.

The polygraph examiner faces two major challenges when testing subjects from another culture. The first goal is to maximize the validity of the exam itself by assuring that cultural differences do not bias the basic procedure. Directing the examinee's psychological set well enough to obtain interpretable charts requires a basic familiarity with that examinee's culture. For example, Russell's (1989) circumplex model of emotions indicated that anger is common to all cultures and (at least in the ones he studied) is close to the emotion of fear physiologically and psychologically. The examiner's task is, therefore, to maximize fear of detection while avoiding treading on the subject's culturally-determined values, thereby possibly producing invalid results.

DoDPI Research, Fort McClellan, Alabama, has undertaken an effort to determine how cultural factors may (or may not) impinge on physiological responses of individuals during polygraph examinations.

[1].This manuscript was completed in 1990 as a final report to the Department of Defense Polygraph Institute.

The existence of such impingements can obviously have a direct influence on the validity of polygraph administrations and, in turn, enhance or detract the detectability of deception through clinical judgments and affect the inducement of confessions to crimes before, during, or after a polygraph examination. In an effort to determine such factors DoDPI, by the way of Requests for proposals (RFP's), awarded a contract to The Bass Group (TBG) of Pensacola, Florida. The contract divides the effort into three phases. Phase I entails an extensive literature review seeking out research related to each of the following general factors:

(1) Truth, lying, shame and guilt.

(2) The belief in the validity or efficacy of detection of deception.

(3) Cultural tradition of autonomic control.

(4) Ability to control self- presentation.

(5) Tester/Testee interactions and critical examiner behavior.

Phase II of this research effort will consist of TBG's designing a questionnaire reflecting the results of Phase I. It is DoDPI's intention to administer this questionnaire to polygraph examiners who have experience and expertise testing foreign nationals. Finally, in Phase III, TBG will develop a written position paper suggesting a prioritized program of research relating to the whole issue of cross-cultural factors and physiological responses to polygraph testing. The present report contains the results of Phase I.

Definition of Culture

At the outset, DoDPI and TBG grappled with the definition of culture in an attempt to provide an operational definition. It became clear that providing a definition of culture had incurred the time and energy of several respected authors.

Tylor (1877) is recognized as the first to use the word culture in English as generally accepted by most anthropologists and sociologists. Tylor defined culture as "that complex whole which includes knowledge, belief, art, morals, law, custom, and any other capabilities and habits acquired by man as a member of society" (p. 1).

Linton (1945) defined culture as "the configuration of learned behavior and results of behavior whose component elements are shared and transmitted by the members of a particular society" (p. 1). Barnow (1985) makes the point that several authorities while liking the integration stressed in this definition objected to the inclusion of "results of behavior" in the definition.

Bockner (1982) points out that Kroeber and Kluckhohn (1952) reviewed over 150 definitions of culture, all of them plausible, and that inordinate amounts of time could be dedicated to exploring their relative merits without resolving the issue. Brislen (1983) in discussing definitions of cultures identifies definition by Krober and Kluckhohn as being the most widely accepted which included elements such as "patterns, explicit and implicit, of and for behavior acquired and transmitted by symbols, constituting the distinctive achievements of human groups . . . (and) ideas and their attached values " (P.367). Brislin echoes Bockner when he states that there is no one definition of culture, which is widely accepted.

Triandis (1980) states that:

> Culture is one of those motions that are ever present in the work of social scientists, but one that has been defined in so many different ways that no consensus has emerged . . . An important aspect of culture is that it does have structure and it is patterned . . . Cross-cultural refers to comparisons of cultures. This immediately raises the question of what units to be compared. Some cross-cultural psychologists compare cultural units. Others compare larger units, including nations. There are conceptual problems in comparing nations, races, religions, or any other large groupings of people since there is tremendous heterogeneity among the people in such units. In spite of this obvious disadvantage, it is sometimes necessary and fruitful to compare larger units (p.2).

It became evident to the authors of the present study that it was not feasible to adopt a classic anthropologistic definition of culture. To do so, for instance, would not have allowed for studies involving West Germany or East Germany to be included in the study. Examination of the psychological literature clearly suggested that the vast majority of studies entitled "cross-cultural" involved comparisons across countries. For purposes of this study the review of the literature includes comparisons between and among countries or possessions fully recognizing that within any one country or possession there may exist several cultures. The following were considered for inclusion in this review of the literature:

1. Argentina
2. Austria
3. Brazil
4. Chile
5. Columbia
6. France
7. W. Germany
8. E. Germany
9. Greece
10. Hong Kong
11. Israel (Jews)
12. Italy
13. Japan
14. Mexico
15. New Zealand
16. Pakistan
17. Peru
18. Philippines
19. Portugal
20. South Africa
21. Singapore
22. Soviet Union
23. Spain
24. Taiwan
25. Thailand
26. Turkey
27. USA
28. Venezuela
29. Yugoslavia

Method

The search strategy for this literature review consisted of three major approaches: (1) a multi-database strategy that used input terms "culture" or "cultural factor" AND "polygraph" or "lie detector", (2) combining "cultural factors" or "cross-cultural" terms with each of approximately 25 major descriptors or terms deemed relevant to factors via, physiological, psychological, sociological that might influence the validity of a polygraph exam, and (3) utilizing a similar search strategy as noted in (2) with each of 29 countries or possessions. The intent was to obtain a broad, comprehensive search using a variety of specific culturally related terms across several key databases.

There were several reasons for adopting the above approaches. The strategy of combining "cross-cultural factors" with "polygraph" unfortunately produced disappointing retrieval results. Apparently, there is a dearth of research on this topic in the open literature. Thus, it was decided to select "key terms" that tap on cross-cultural factors, which may have an influence on polygraph validity. These "key terms" were determined by selectively choosing major areas that have a demonstrated influence on cross-cultural research in general. Various textbooks on polygraph/lie detector and cross-cultural research were selected from books in print and their contents studied.

In addition, major books in sociology and psychology with an emphasis on cultural factors were pursued for major areas and topics that have shown to have inter-cultural influences. From this investigative approach a number of major areas and topics were selected as descriptors and identifiers (input) terms for the multi-database strategy and computer searches.

Databases

The following databases were searched by computer: (1) Biosis Previews, (2) Embase, (3) Legal Resource index, (4) Medline, (5) Mental Health Abstracts, (6) National Criminal Justice Reference Service (NCJRS), (7) National Technical Information Service (NTIS), (8) Public Affairs Information service (PAIS International), (9) PsycINFO (Psychological Abstracts), (10) Social Scisearch, and (11) Sociological Abstracts. A description of each of these databases can be found in Appendix A.

Search Terms

The following are the search terms employed in database searches:

1. Physiological Reactivity
2. Galvonic Skin Response (GSR)

3. Autonomic Reactivity
4. Blood Pressure
5. Socialization
6. Acculturation
7. Intelligence (IQ)
8. Birth order/Family size
9. Lying
10. Truth/Truthfulness
11. Guilt
12. Self-Disclosure
13. Values
14. Sexual mores
15. Stress
16. Anxiety
17. Emotions
18. Deceit
19. Attribution of Blame
20. Paranoia
21. Discipline
22. Defense Mechanisms
23. Fear
24. Interrogation
25. Suggestibility
26. Confession
27. Deception

In addition to the above-listed computer searches the authors conducted a two-day on-site visit to the University of Florida library in Gainesville, Florida, in order to access the Human Relations Area Files (HRAF). HRAF is a private, nonprofit educational institution which has compiled achives consisting of over 800,000 pages of text on over 330 different cultures (Levison, 1989). A thesaurus titled *The Outline of Cultural Materials* contains general and specific descriptors, which can be used to assist researchers in finding particular topics. Research noted by Levinson (1989) indicates that using the subject-category codes provided in *The Outline of Cultural Materials* will allow most researchers to find 90 percent or more of available archival materials on any given subject.

In an effort to supplement data that exists in the open literature and archives TBG conducted several phone interviews with noted authors who are presently involved in cross-cultural research. While this effort, in general, produced minimal results, it did on occasion prove fruitful.

Results and Conclusions

This section is divided into two parts, via, (1) a review of the literature and (2) a rational deductive approach based upon Hofstede's (1980) theory of cultural differentiation.

Review of the Literature

It should be stated at the outset that there exists a paucity of data, which directly relates to the five factors outlined in the Introduction. It quickly became evident that the cross-cultural literature, especially the anthropological literature, is replete with anecdotes, personal observations, and judgments. While such data can be useful there is some hesitancy to allow them to serve as the foundation for establishing procedural standards for testing such as in polygraph examinations.

Truth, Lying, Shame and Guilt

It was rather surprising to the authors that there exists relatively little research in the open cross-cultural literature on the subjects of truth, lying and shame. One is tempted to speculate that these topics are laden with such negative stigma that researchers have purposefully avoided their investigation.

William R. Johnson (1987) was a former military intelligence agent during World War II and worked for the Central Intelligence Agency for 30 years. In an article entitled "Ambivalent Polygraph" he made a number of observations about testing in cross-cultural situations:

> The polygraph operator working overseas learns to modify this theory somewhat. He finds that it applies uniformly to the sexual conscientiousness of Northern Europeans, natives of the British Isles, and Americans, who share a common guilt-culture, but less so to Latin Americans, Southern Europeans, and Middle-Eastern Moslems, and it applies hardly at all to Southeast Asian non-Christians . . . Fortunately, however, sex is not everything. In most cultures, speaking truth is a virtue and lying is a vice. Even with those subjects whose

> culture has conditioned them to say, out of politeness, what will please their host, the experienced operator can create an interpersonal situation in which the subject feels obligated to speak the truth to the interrogator whether or not it is polite (Johnson, 1987, p. 300).

However, truth while being considered a virtue in most countries can vary in terms of its importance. For instance Howard, Shudo, and Umeshima (1983) comparing Japanese and American business managers on the Rokeach Values Survey found the Japanese to score significantly higher on the Honesty scale. Russians feel duty to be the highest of motives the one " . . . to which all others are to be subordinated" (Bauer, 1952, p. 140). The Russians value truth among their fellow citizens but will unhesitatingly lie if they perceive doing so as a duty to the state. Mead (1951) points out that "Communists will try to compensate for their professional severity and are often more honorable in their personal relations than people who effect ethics" (P. 32). Confucian teachings emphasize moral ideals and place the virtue of social justice above any considerations of utility (Wu, 1967). Danton (1938) notes that many foreign visitors to China feel that the Chinese do not have a conception of truth and will lie under any circumstances.

Dorothy Lee (1953), a native Greek who conducted ethnographic research for UNESCO, states that honesty and obedience are important and are taught to children from the beginning, but this is taught in the framework of personal relationships with family and immediate friends and not to other groups.

Dickson (1949) observes that the cultural tradition of Rojal Kahoir suggests that lying to prevent problems between people is acceptable in the Arab culture. In the case of deceptive subjects, the polygraph examiner may take the approach during the interrogation that he or she understands the examinee was attempting to serve as a Raja Kahair but that telling the complete truth is the more appropriate way to be a good person.

Discussion on the topic of shame and guilt emanates more from observations and subjective judgments rather than experimental research. Mead (1951) in discussing the conflict between the severity of the Russian's allegiance to the state and his inner conscience states "calm consciousness as a result of awareness of duties fulfilled gives man immense joy and the one who experiences torments of conscience from bad actions bridges a social duty, feels terribly oppressed" (p. 47).

There is indirect evidence in the anthropological literature to support the notion that the Japanese could experience both guilt and shame sufficient to arouse physiological indices of deception (DeVos, 1960, 1968).

Gorkin's (1986) article on counter-transference in cross-cultural psychotherapy is worth noting. In studying Jewish psychotherapists who were treating Arabic patients, the emphasis was on how the therapists felt toward their patients because of their ethnic and religious differences. Counter-transference addresses the issues, which arise when the therapist projects his or her values and feelings onto the patient; or when the patient's appearance, characteristics, or values trigger strong feelings in the therapist. Gorkin found that the therapists (in one-up positions due to their status as therapists) tended to feel guilt about Israeli-Arab relations and let that guilt affect their perceptions of Arabs. That is, although "one-up" due to their official status, the therapists tended to feel "one-down" due to guilt. Also, their curiosity about Arab culture often led them to explore details about the patient's life-style rather than focus on her or his psychological problems.

Gorkin's ideas are important for polygraph examiners because they illustrate potential pitfalls. Becoming too concerned about a subject's culture can distract the examiner from observing important nonverbal behavior. It is also necessary to be attuned to positive and negative feelings about other cultures so that such attitudes do not lead the examiner into projecting his or her feelings into the interpretation of polygraph charts.

Belief in The Validity/Efficacy of Detection of Deception

No cross-cultural studies exist, to the knowledge of the authors, which directly

assess a culture's belief values in terms of detecting deception. However, Barland (1988) reports that there are six countries with a major polygraph capability, via, in the order of estimated number of examiners: (1) USA, (2) Canada, (3) Japan, (4) Turkey, (5) South Korea, and (6) Israel. Barland also estimates there to be at least thirty other countries with one or more polygraph examiners. It is probably safe to assume that the polygraph is employed by the Soviet Union and perhaps other communist countries.

However, use of the polygraph is not a perfect criterion for determining a culture's belief in the validity of measures of detection. In some countries such as Germany the technique was declared illegal for use by both the courts and the law enforcement agencies not because it was considered to be invalid but rather because it was seen as an encroachment upon the freedom of the individual (Kaginiec, 1956). In Holland the instrument is rendered useless since the examinee is not required to render any assistance nor ask any questions (Meyer, 1961).

Cultural Tradition of Autonomic Control

Some cross-cultural experimental studies investigating autonomic control do exist. Guthrie (1975) states that despite a common autonomic nervous system culturally different groups of people have different psychophysiological reactions to stressful situations. However, this finding should not be surprising since stress is a function of the perception more so than the function of the stressor. Such perceptions are clearly modified and molded by the values, attitudes and mores of each culture.

Kugelmass and Lieblich (1968) reported that subjects of Near Eastern origin tended to have lower pulse rates, higher basic skin conductance, and lower relevant GSR reactivity. Japanese were found to have higher skin conductance than Americans in an experimental study involving the watching of a stressful film (Lazarus, Tomita, Opton, & Kodama, 1966). Ohnishi, Matsuno, Arasuna, and Suzuki (1976) found that detection rates for respiration were only accurate 46 percent of the time with Japanese subjects, while electrodermal measures had a 72 percent accuracy rate.

Waid and Orme (1981) compared skin conductance responses using an intrusive biographical interview among college students who were English, German, Irish, Italian, Jewish, and Scottish. The findings indicated significantly smaller EDRs among the Irish than the other five groups. In addition, the data suggested that people are less aroused physiologically if the interview is conducted by an individual of their own background as opposed to an interview conducted by a person of a different background.

Ability to Control Self-Presentation

Mead (1951) observed that it is not unusual for Russians to change from extreme uncooperativeness to some semblance of cooperativeness. Mead compares this deportment with that of the Japanese prisoner and a Pole. In this comparison Mead states that the Japanese prisoner is likely to succumb to pressure within a few hours after being taken prisoner. While a Pole is apt to "remain actively intransigent despite drastic changes in circumstances" (p. 37).

Dicks (1952) makes a similar observation of Russians stating that a "Russian may be rated by someone against whom he is powerless suddenly throw up his hands and say 'shoot me then if you like, what do I care?" Dicks suggests that such outbursts are an attempt to arouse guilt in the aggressor and appeal to his mercy. Anisimov (1951) makes the observation that the Soviet citizen will assume at the slightest provocation on a foreigner's part a mask of arrogance, a supreme contempt for what he will describe as the "outward" and "superficial" civilization of the West. Russians are apparently current in the state-of-the-art of self-control over mental states. Gabndreyeva and Peesahhov (1982) describes an elective course, offered at the University of Keegan designed for students who suffer from excessive shyness, test anxiety and similar problems. The training is in the form of autogenic training including self-suggestion, visualization, relaxation, and breathing exercises.

In a cross-cultural study of test anxiety in Iranian and Indian students, Sharma, Parnian, Speilberger (1983) found that Iranian students had higher anxiety as measured by the Test Anxiety Inventory than did the Indian students. The authors make the observation

that tricultural differences in the test anxiety levels of comparable students in Iran, India, and the United States indicate greater test anxiety in Eastern cultures.

Feldman (1983) believes there are significant cultural differences in the nature of nonverbal expressivety between Koreans and Americans with the former having a greater degree of control of facial nonverbal behavior. In a study conducted by Ekman, Friesen and Ellsworth (1972) comparing college students in the United States and Japan, it was observed that both groups of students essentially manifested the same negative facial expressions while watching the stress film. However, in a subsequent interview with a member of their own culture, the Japanese displayed happier or more impassive faces in describing the stress film. The authors concluded that cultural rules can override universal experience.

Eysenck (1982) conducted an extensive study on cultural differences with regard to mean levels of anxiety, introversion-extroversion, and neuroticism. Eysenck's goal was to study the relationship of personality factors such as introversion-extroversion and blood type. He gathered data from Hofstede's study (1980) of over 70,000 subjects. Hofstede conducted an extensive survey of people in over 40 cultures, and one of the questions he asked them was as follows: "How often do you feel nervous or tense at work?" Eysenck gathered these data, as well as data on blood types, in hopes of developing theories about the importance of genetics on personality differences between nations. Of special note to this study is that different countries were then lumped into the categories of high anxiety, average anxiety, and low anxiety. Countries that were found to fall in the high anxiety category were as follows: Japan, Greece, Belgium, Argentina, Colombia, Yugoslavia, and Taiwan.

Countries in the average range of anxiety on Hofstede's question were as follows: Italy, Spain, France, Turkey, West Germany, South Africa, and Canada. Countries found to fall in the low anxiety category were as follows: Australia, United states, Ireland, Great Britain, New Zealand, Sweden, and Denmark.

Further looking at the introversion-extroversion category, Eysenck labeled the following countries as falling into the extraverted category: Australia, Canada, Greece, India, Poland, South Africa, Sweden, United Kingdom, United States, and Italy. Countries falling in the introverted category were as follows: Egypt, France, West Germany, Iran, Japan, Turkey, and Yugoslavia.

Russell, Lewicka and Niit (1989), and Epkman and Friesen (1986) address through the use of research the question as to whether or not there is a range of emotions common to all cultures. Ekman and Friesen studied the use and interpretation of facial expressions across cultures. They took photographs of different individuals with presumably certain emotions being experienced and showed those photos to subjects in a number of different cultures to determine how accurately they could pick out the designated emotion. Ekman, et. al. found that contempt was recognized accurately by subjects from Estonia S. S. R., Germany, Greece, Hong Kong, Italy, Scotland, Turkey, United States, and West Sumatra. Ekman (1987) also found that there was considerable agreement among subjects from Estonia, West Germany, Greece, Hong Kong, Italy, Japan, Scotland, Sumatra, Turkey, and the United States in recognizing facial expression of the following emotions: happiness, surprise, sadness, fear, disgust, and anger. Wolfgang and Cohen (1988) developed a technique called the Wolfgang Interracial Facial Expression Test (WIFET) and found that it was, indeed, harder to read facial expressions when the expression was on the face of a person of another race. However, Wolfgang and Cohen noted that people can be trained to do so and stressed the need for a WIFET for each major culture to be used in future research.

Russell, Lewicka, and Niit (1989) developed what they term a circumplex model of affect. They used a multi-dimensional scaling of pair-wise similarity scores to develop his research. The bipolar dimensions of pleasure-displeasure and arousal-sleepiness were used, and subjects were asked to sort "feeling" related concepts in a pairing manner. The data were placed into the bi-polar dimensional model. Subjects were from the language groups of Estonia S.S.R., Greece, Poland, and China. Russell et al. found that if

one plotted the different "feeling" concepts on a two dimensional graph where arousal and sleeping were the vertical line, and pleasure and displeasure were the horizontal line, that most of the "feeling" concepts (each represented by a dot) would tend to cluster in a circular manner. For example, the common concepts of afraid and angry (each represented by separate dots) tend to cluster in the upper-left quadrant formed by arousal and displeasure. The "feeling" concept of relaxed clustered in the lower-right quadrant designated by pleasure and sleeping. What this means is that people tend to use the word relaxed when they are describing sensations that are pleasurable and involve a lower state of physiological arousal. Likewise, when individuals use words such as afraid and angry, they tend to be describing a state, which is a combination of physiological arousal, and the experience of displeasure or annoyance. Feelings of depression tend to show up in a quadrant between displeasure and sleepy, which indicates that the concept of depression taps into lower physiological arousal which is displeasurable in nature.

What is interesting in Russell's study is that he found that if one placed the concepts on the graph, they tend to form a circular type of pattern and this circular pattern was essentially the same for Estonians, Greeks, Polish, and Chinese groups. The presence of such uniformity in the experience of emotions would suggest that polygraph examinations which tap those emotions should theoretically be possible in many cultures.

Tester/Testee Interactions and Critical Examiner Behaviors

Of all the topics reviewed in this study none are as widely represented in the research literature as that of tester/testee interactions and critical examiner behaviors.

Triandis and Brislin (1988) and Brislin (1989) emphasize that the concept of individualism and collectivism are important topics for coverage in cross-cultural training programs. In defining the two concepts the authors state that "individualism is characterized by the subordination of a group's goals to a person's own goals" while "collectivism is characterized by individuals subordinating their personal goals to the goals of some collectives" (p. 269). Triandis, et. al. suggest several principles that should be considered when training collectivists to interact with individualists and when training individualists to interact with collectivists. The authors of this study believe that the latter could be useful for American polygraph examiners testing individuals from a collectivistic culture. For this reason these principles are presented in Appendix B.

Triandis (1985) points out that in general, collectivist cultures almost always assume that in-group members in authority do the right thing but out-group authorities are viewed with suspicion. In the case of polygraph testing of a member of a collectivist society it would appear to be wise to have the examiner be of the same culture and older in age. In the examination of a member of an individualistic culture such considerations would not be important.

Triandis associates the following cultures as being collectivistic: (1) Southern Europe, (2) Northern Europe (certain cultures that have retained traditional element(s), (3) South America, and (4) East Asia. Triandis points out that most Western cultures are individualistic, especially the United States, Britain and the former British Empire.

Samover, Porter and Jain (1985) see the Soviet society as placing strong emphasis upon the group or collective, deriving its conception of the relationship between the individual and the state from communist doctrine.

Several studies have conducted cross-cultural research regarding body language. Watson and Graves (1966) comparing Arabs and Americans found that Arabs were more direct in face-to-face orientation, maintained less distance from one another and touched more. The Arab subjects also had more direct visual contact and spoke louder than Americans. Patai (1973) states that in dealing with the Arab, exaggeration should not be taken literally, but only as a technique used for effect. Watson (1970) studied the gaze of pairs of students from different countries. The highest level of gaze was manifested by Arabs and Latin Americans, the lowest by northern Europeans. Argyle, Furnham and Graham (1982) make the

observation that too much eye contact is considered threatening to the Japanese.

Polygraph examiners would be well-advised to sit directly in front and close to Arabic and Latin examinees, rather than conduct the pre-test interview from behind a desk. Examinees should be prepared for occasionally being touched while establishing rapport in the pre-test interview and not recoil or display negative emotions should the examinee do so. During the pre-test interview with oriental (especially Japanese) subjects, too much eye contact might over-stimulate the subject and result in over-reactive (and hard to interpret) charts. For deceptive Japanese subjects, however, a long, direct eye-gaze during the post-test interrogation might increase anxiety and facilitate obtaining a confession.

Montague (1971) observed that people who speak Latin derived languages are more contact-oriented than those who speak Anglo-Saxon derived languages. That is, people from places such as Italy, France, South America and Mexico may prefer more contact in social situations by coming closer, touching more, and using expressive gestures more than English, Canadians, and Americans. Chan (1979) found that the Chinese express anger and disgust by narrowing the eyes, the reverse of that found in the United States.

Cleveland, Mangone, and Adams (1960) state that in Asian countries the word "no" is rarely used. Alternatively, "yes" can mean "no" or "perhaps". They also observe that Arabs have a "run in" period of informal chat for approximately half an hour before getting down to business.

Examiners should be cautious that Japanese subjects may appear to be confessing when in fact they are just being polite. Many examiners make up a "saving face" kind of story to help deceptive subjects minimize their guilt and more readily confess. As the examiner carefully tries to sell the subject on the confession story, the oriental examinee might sit quietly and politely nod, thereby appearing to agree with the examiner (i.e., confess). To prevent such mistakes, attempts to obtain confessions should never be made until after the polygraph testing of subjects from collectivist cultures. This is probably already DoDPI doctrine, but deserves being underscored.

Several studies have investigated distance, seating arrangements, positioning and territoriality in general. Lomrantz (1976) found interaction distances among members of Mediterranean cultures such as Greeks, and Southern Italians preferred closer distances than Northern Europeans. Argentinean students sought the greatest distance from strangers, placing them almost nine times as far apart as friends. The Iraqis, on the other hand, preferred the smallest interaction distances and made little distinction based on relationship. Lomrantz also found in a study of student immigrants from Argentina, Iraq, and Russia found that all three cultures preferred closer distances with friends than strangers and with a fellow countryman than with an Israeli.

A few studies have reviewed the subject of positioning, distance, and arrangement across cultures. Pakistani subjects viewed opposite seating as more distant than did the other groups. Watson and Graves (1966) confirmed earlier observations that Arabs, compared to Americans and Western Europeans, stand much closer and also adopt a more directly facing orientation. Cline and Puhl (1984) in a comparison of desired seating arrangements found that Chinese preferred side seating compared to U.S. subjects. The authors felt that the corner seating preferred by U.S. subjects would be viewed as aggressive by Chinese subjects.

Argyle (1982) makes several observations about intercultural communication. In comparing Americans with Mexicans, he notes that Mexicans regard openness as a form of weakness or treachery and are very protective of allowing the outside world to penetrate their thoughts. In discussing bodily contact Argyle states that cultures vary significantly. "Contact" cultures include Arab, Latin American, South European, and some African cultures. In "non-contact" cultures, bodily contact is confined to the family. Exceptions include greeting, parting, and professional behavior, for example hair stylists, physicians, etc. Considerable anxiety can be created from bodily contact outside these settings.

In light of the limited information about the Soviet Union it might be useful for the purposes of this study to include some observations made by Dicks (1952). He notes that the Russian frequently appears to be the innocently accused and persecuted whose aggression is purely defensive. Anybody that looks at all safe tends to evoke the need to share, which Dicks believes may explain in part the Russians falling easy victims to the secret informing systems.

A Rational-Deductive Approach

The great majority of research discussed so far focuses on the individual nuances of particular cultures. Very little is published synthesizing and comparing large numbers of cultures along universal dimensions. This results in a fragmentation of knowledge about the fundamental similarities and differences among cultures.

Hofstede (1980) published a monumental work, integrating voluminous data on forty different countries. In this work he describes a research project on employees of a large, international conglomerate. The employees (in 40 countries) were surveyed extensively on work-related values using a questionnaire, which correlated with well-known instruments such as Schultz's FIRO-B, England's Personal Values Questionnaire, and The Allport-Vernon-Lindzey Study of Values. The sample included unskilled, skilled, clerical, nonprofessional sales, technicians, professional and managerial workers.

The questionnaire was administered between 1967 and 1973 to 88,000 respondents and is easily one of the largest cross-cultural databases available today. Over 50 occupational groups in 40 countries were surveyed on overall job satisfaction, perceptions of stress, personal goals, attitudes, and beliefs and an extensive factor analysis carried out.

Hofstede's factor analysis resulted in four factors or dimensions, which he labeled Power Distance (PDI), Uncertainty Avoidance (UAI), Individualism (IDV), and Masculinity (MAS). Hofstede notes that the type of supervision preferred by employees, their willingness to disagree with a boss, and how they perceive their supervisor's style of decision making constitute the dimension of Power Distance. Respondents who score high on the PDI scale tend to prefer a hierarchy in social relationships in which everyone has his/her place, there is a clear demarcation between superiors and subordinates, power holders are entitled to privileges, other people can rarely be trusted, and power holders should attempt to look as powerful as possible.

Countries scoring high on this index include the Philippines, Mexico, Venezuela, and India, with Japan being just above the mean and West Germany (surprisingly) being below the mean. Changes in German society as a result of World War II may account for this finding.

Low PDI scores correlate with the following societal norms: no clear demarcation between superiors and subordinates, a belief in equal rights for all, harmony among fellows is prized, trust is an important value, and powerful people should attempt to look less powerful than they are. Countries scoring low on PDI include Austria, and Israel. Table 1 rank orders the ten countries highest in PDI and Table 2 rank orders the ten countries lowest in PDI. In addition, Tables 1 and 2 show their rankings on UAI, IDV, and MAS.

Because of the extensive sample size and number of countries involved, Hofstede's four dimensions will serve as the organizing structure for the remainder of this section. He used the scores on these dimensions to group 40 countries into similar categories, with each category representing similar clusters of traits and norms. These categories were as follows: more Developed Asian, Less Developed Asian, Near Eastern, Germanic, Anglo, Nordic, More Developed Latin, and Less Developed Latin. These categories will serve as the titles for subsequent sections of this paper. However, Anglo and Nordic countries are excluded and a new category representing the Middle Eastern countries will be added (although Hofstede's research does not cover this grouping).

Table 1
Ten Countries Highest in Power Distance (PDI), with Relative Rankings in Uncertainty Avoidance (UAI), Individualism (IDV), and Masculinity (MAS)*

COUNTRY	PDI	UAI	IDV	MAS
Philippines	1	32	28	10
Mexico	2	11	29	6
Venezuela	3	14	39	3
India	4	33	21	19
Yugoslavia	5	5	31	36
Singapore	6	39	33	24
Brazil	7	15	25	23
Hong Kong	8	36	31	17
France	9	6	11	29
Colombia	10	13	38	11

Note. * Source: Hofstede, G. (1980)

Table 2
Ten Countries Lowest in Power Distance (PDI), with Relative Rankings in Uncertainty Avoidance (UAI), Individualism (IDV), and Masculinity (MAS)*

COUNTRY	PDI	UAI	IDV	MAS
Great Britain	30	34	3	8
Switzerland	31	24	14	5
Finland	32	23	17	35
Norway	33	27	13	38
Sweden	34	37	10	39
Ireland	35	35	12	7
New Zealand	36	29	6	15
Denmark	37	38	9	36
Israel	38	12	19	25
Austria	39	18	18	2

Note. * Source: Hofstede, G. (1980)

Hofstede points out that we all live with uncertainty about the future and our place in it and that we vary in the anxiety associated with this uncertainty. In organizations, some people prefer a loosely structured work environment and tolerate ambiguity well. They tend to score low on Hofstede's uncertainty Avoidance Index. High scorers tend to prefer security and structure in their lives. High anxiety and stress are common, as is a belief that "time is money", intolerance is common, and conversation-law-order are stressed. Japan, Germany, Greece, and Peru score about the mean on UAI. The United States scores below the mean.

The third dimension (Individualism) reflects differences in desire for personal time and needs versus loyal service to the company/organization/country. High scores on IDV reflect an emphasis on individual achievement, autonomy, and the self. The United States, Australia, Britain, and Canada score high on this dimension. Low scorers (such as Iran, Mexico, and Chile) emphasize a collective orientation, group belonging, and the source of security as deriving from the social grouping.

Masculinity is the final factor isolated by Hofstede. High scores reflect ego-oriented, assertive approaches to life and low scores reflect socially-oriented, nurturing approaches.

More Developed Asian-Japan

Hofstede's research places Japan in a category which is noted by high masculinity scores, high uncertainty avoidance, and medium individualism and power distance. one would, therefore, presume that many Japanese would prefer a relationship with supervisors (and potentially polygraph examiners) which were characterized by an emphasis on some of the following characteristics: an emphasis on obedience, high value placed on conformity, potentially authoritarian attitudes, and a preference for an autocratic style of decision making by people who have power over him or her. Hofstede's theory would predict that the Japanese would be fearful of disagreeing with their employers (and potentially polygraph examiners or police officials). Also predicted would be a belief that time is money, a strong sense of nationalism, and a strong belief in experts and their knowledge.

Near Eastern Cultures

Near Eastern cultures were characterized by Hofstede as being high in Power Distance, high in uncertainty avoidance, low in individualism, and medium in masculinity. He listed four countries as falling into this category: Greece, Iran, Turkey, and Yugoslavia.

Hofstede includes Yugoslavia in the Near Eastern category because of the similarities along his four dimensions. He lists Yugoslavia as being high in power or distance, uncertainty avoidance, and low in the other two dimensions. However, Hofstede indicated that he did not have a representative sample of Yugoslavian workers and his insights on Yugoslavia may be of limited generalized ability.

Less Developed Asian

The less developed Asian nations which were noted by Hofstede (1980) were as follows: Pakistan, Taiwan, Thailand, Hong Kong, India, Philippines, and Singapore. For purposes of this paper China and Korea are also included. Hofstede characterizes most of the less developed Asian countries as being high in power distance, low to medium in uncertainty avoidance, low in individualism, and medium in masculinity. The overall cultural stereotype that one might expect from such a constellation would be a strong emphasis on adherence to authority figures or respect for authority figures, less anxiety and internalized stress than one would expect in a country such as Japan, a strong interest in collective and cooperative functions, and somewhat above average emphasis upon achievement and assertiveness.

It is interesting to note that Singapore, Hong Kong, India, and Philippines all tend to have fairly high scores on masculinity and fairly low scores on uncertainty avoidance. Hofstede suggests that such countries could be conceptualized as "masculine risk takers" (p. 324) and notes that these are formally colonies of the United States or United Kingdom. Hofstede states that individuals from such countries tend to focus on meeting their ego needs through achievement and hard work, and tend to be more motivated by a hope of success than by a fear of failure.

Malcolm (1951) served as a Justice in the Supreme Court of the Philippines and notes in his analysis of Philippine culture that the Chinese minority have a very strong reputation for honesty and integrity. It would probably be safe to suggest that there are likely differences among the Chinese who live in mainland China, low zone Taiwan, and those in areas such as Hong Kong. For example, Hofstede found that Taiwanese tended to be fairly high on uncertainty avoidance, and individuals from Hong Kong tended to be fairly low on uncertainty avoidance. This would suggest that a hope of success is a stronger motivator for citizens in Hong Kong and that fear of failure may be a somewhat stronger motivator for citizens in Taiwan. To the extent that low uncertainty avoidance in a culture predicts low levels of anxiety in individuals from that culture, it would be expected that subjects from Singapore, Hong Kong, India, and the Philippines would be low in anxiety and would need more stimulation in the pretest interview in order to obtain good polygraph charts. Individuals from Thailand, Taiwan, and Pakistan might require less stimulation and more reassurance in order to get readable polygraph charts. Subjects from these last countries might be well motivated through fear of failure or punishment to confess when they are deceptive.

Middle Eastern

Middle Eastern countries would include Syria, Iraq, Saudi Arabia, Egypt, Libya, Jordan, and Lebanon. Israel was considered by Hofstede in his work to have a cluster of dimensions more similar to the Germanic countries and is considered to be part of that cluster for our discussion.

One of the most unifying factors in looking at the cluster of countries in the Middle East is that they are primarily Arabic speaking and have a common religious heritage of the Moslem faith.

The Latin Countries

The Latin countries may be divided approximately into the more developed cluster (Belgium, France, Argentina, Brazil, Spain, and Italy) and the less developed Latin countries (Colombia, Mexico, Venezuela, Chile, Peru, and Portugal). Little data are available on the Central American countries.

Using Hofstede's data, these countries generally score high on measures of power distance and uncertainty avoidance. Business and governmental organizations in these countries tend to gravitate toward pyramid shaped bureaucracies. The emphasis is on structure, rules, and conformity in order to avoid anxiety. These countries are predominately Catholic in their religion. Adherence to values such as veracity will be stronger in the presence of the organization (church, government, etc.), but would weaken outside the organization's influence.

Hofstede discussed the development of the conscience (superego) based upon an interaction of uncertainty avoidance and power distance. For example:

> A high UAI score was related to a strong superego. However, in a high PDI environment, this superego will be personified in the form of a powerful person (the father, the leader, the boss). People will be able to blame the powerful people for their ills (a favorite pastime in the Latin countries) and will feel relatively free to sin if the boss isn't looking. In the higher UAI, low PDI countries this escape is not available, and the superego is internalized (Hofstede, 1980, p. 316).

Iran, Thailand, Pakistan, and Taiwan tend to cluster near the Latin countries on these dimensions. Implications for polygraph practice are that if the examiner adheres to religious sanctions for truth and the need for redemption or quotes from the Bible (or Koran as appropriate) the interviews may be facilitated.

The Germanic Countries

Austria, West Germany, Switzerland, and Israel share a constellation of cultural values. They share low scores on power distance and medium to high scores on uncertainty avoidance, masculinity and individualism. Perhaps the placement of Israel in this grouping is to be expected, given the European post-Holocaust emigration.

The Germans, Swiss, and Austrians tend to be more motivated by "ego security" needs and Israel by group solidarity needs. On average, one would expect the needs of the group to have precedence over the needs of the individual. These same four countries tend to cluster into what could be called the "internalized superego" (Hofstede, 1980, p. 316) category. Moderate uncertainty avoidance combined with low power distance suggests strong, internalized conception of right and wrong which operate in the absence of outside coercion.

Organizations in cultures which combine group solidarity, a hard-worker orientation, and a strong conscience tend to run like "well-oiled machines" (Hofstede, 1980, p. 319). One could speculate that soldiers found in Israel and the Germanic countries would be more likely to continue to attempt to accomplish their assigned missions, even if their leaders were killed or missing in the confusion of battle than soldiers from categories such as the Latin grouping.

Gorkin's (1986) study of counter transference noted earlier in this paper also illustrates the tendency toward guilt by Israeli psychotherapists. This illustrates the strong, internalized superego hypothesized by Hofstede for the low PDI, high uncertainty avoidance countries. Perhaps it's not so unusual that psychoanalysis (with its focus on guilt) is still strongest in Germany, Switzerland, Austria and Brazil; or that its founder (viz, Freud) was an Austrian!

It is interesting to note that power distance was relatively low in the Germanic countries. This does not fit a popular conception of the Germanic-speaking peoples as being high disciplined and conforming to authority. Hofstede suggests one possible explanation. Many of his questionnaire items focus on willingness to disagree with the boss.

If one were to look at the potential role that could be played by the polygraph examiner and present it as a dichotomy between two different types of approaches, one could label one end relaxed and the other end professional. The professional examiner would be a person who strongly emphasized rigid adherence to rules and regulations, would be dressed in a semi-military and somewhat rigid-appearing manner, and would project the image of being a "no nonsense, down to business" type of person. The opposite pole of this stereotype would be the very relaxed, laid-back, casually dressed, and imprecise in speech and behavior kind of person. The reader is directed to Table 3, which contains the hypothesis based on Hofstede's four dimensions across 40 countries. In looking at the section under Japan note that under examiner role the term used is professional. The hypothesis is that a polygraph examiner working with a Japanese subject would have the most positive effect on that subject if he or she assumed a role approximating the professional end of the continuum. The professional behaviors would consist of traits already noted in the expectation that the Japanese subject would be more comfortable with such an individual and be more likely to respect the professional approach.

Note that the next column in Table 3 is labeled Polygraph Validity Approach. This column is designed to generate hypotheses about how the polygraph examiner should best approach the examinee regarding the issue of a polygraph's validity. There is considerable anecdotal information in the literature and debate about the best way to introduce the technique and technology to the examinees. One approach stresses the simple explanation that the polygraph is valid, almost to the point of being infallible and that the examiner will make no mistakes. This is designed to convince the examinee that no mistakes will be made and allow the truthful subject to relax and the deceptive subject to experience sufficient fear to be detected. Some examiners stress the importance of stimulating subjects with cards, tests, and other techniques for convincing them that the polygraph is very valid. In the topology noted on Table 3, infallibility is used in the column in which a country or culture has been listed by Hofstede as being high in uncertainty avoidance. Individuals who are high in uncertainty avoidance would be anxious people with strong development of conscience and probably would not need to be stimulated to a great extent. On the contrary, such people might be overly anxious and in need of calming.

Table 3
Potential International Approaches Using Hofstede's (1980) Dimensions Across Cultures

	Examiner Role		Polygraph Validity Approach		Individualism		Masculinity	
Country	Relaxed (Low PDI)	Professional (High PDI)	Stimulation (Low UAI)	Infallibility (High UAI)	Self Blame (Low IDV)	Other (High IDV)	Ego (High MAS)	Social Goals (Low MAS)
Argentina	√		√			√	√	
Austrailia	√		√			√	√	
Austria	√			√	√		√	
Belgium		√	√		√		√	
Brazil		√		√		√		√
Canada	√		√		√			√
Chile		√	√			√		√
Colombia		√		√		√	√	
Denmark	√		√		√			√
Finland	√		√		√			√
France		√		√	√			√
Great Britain	√		√		√		√	
West Germany	√			√	√		√	
Greece		√	√			√	√	
Hong Kong		√	√			√	√	
India		√	√		√		√	
Iran		√		√		√		√
Ireland	√		√		√		√	
Israel (Jews)	√			√	√			√
Italy	√			√	√		√	
Japan		√		√	√		√	
Mexico		√		√		√	√	
Netherlands	√		√		√			√
Norway	√		√		√			√
New Zealand	√		√		√		√	
Pakistan		√		√		√		√
Peru		√	√			√		√
Philippines		√		√		√	√	
Portugal		√	√			√		√
South Africa	√		√		√		√	
Singapore		√	√			√		√
Spain		√	√		√			√
Sweden	√		√		√			√
Switzerland	√		√		√		√	
Taiwan		√		√		√		√
Thailand		√		√		√		√
Turkey		√		√		√		√
USA	√		√		√			√
Venezuela		√		√		√	√	
Yugoslavia	√		√			√		√

An entry of infallibility in Table 3 indicates that one would expect individuals from these cultures to be similar and a simple explanation by the examiner stating that the polygraph was infallible and valid would be sufficient. In countries listed by Hofstede as low in uncertainty avoidance, the term stimulation is used in Table 3 to suggest the possibility that such subjects might need extra stimulation to ensure that they will respond physiologically during the test.

Hofstede's data suggest that the Japanese come from a culture high in uncertainty avoidance, and one would hypothesize that the infallibility approach would be appropriate. The third column of Table 3 is a listing based upon the individualism score for different countries. It is designed to generate a hypothesis about how individuals who are deceptive might respond to different interrogational ploys. Column 4 of Table 3, which is based upon the masculinity dimension, is similarly designed. Countries in which the emphasis is upon high individualism would be expected to be oriented toward the self, toward furtherance of the examinee's self-interest. Those scoring in the low range of individualism would be expected to be more other-directed, with an emphasis on others in the environment. Japan tended to score in the medium to high range on individualism and in the high range in masculinity. A high individualism score would presumably result in a self-orientation. A high masculinity score would presumably result in what is noted in Table 3 as an ego approach. Whereas individuals high in masculinity would be expected to be very aggressive in achieving goals for themselves, those at the low end would be expected to be more socially oriented, that is, oriented toward humanistic types of goals. If one looks at Japan and notes a self and ego orientation (based upon high scores in individualism and masculinity), one would expect that the best approach to take in post-test interrogation would be to focus on the ego needs of the individual rather than needs for affiliation.

The United States score on individualism was *higher than* any other of the 40 countries surveyed by Hofstede. The other Anglo countries (Australia, Great Britain, Canada, etc.) followed closely behind the USA. One would, therefore, hypothesize that most polygraph subjects from other countries would view the American examiner as being more of an individualist than they. American emphasis on individual action and responsibility suggests that an American would focus more on his or her contribution to well-being, defending own interests, and a personal philosophy of humankind than people from other cultures. This tendency might result in a stronger tendency to accept self-blame, rather than project blame onto one's organization or associates.

In conclusion, the results of this report strongly suggest that relatively little exists in the literature involving experimental research that directly addresses the cross-cultural factors identified as subject matter for this review. The literature does contain common characteristics across cultures, with considerable variation along a continuum. It also discusses characteristics that are idiosyncratic of a culture or a minimum number of cultures. However, in general, these cultural characteristics are either irrelevant or too general to be useful to a polygraph examiner. It is clear that in order for this void to be corrected there must be a substantial effort in research specifically designed to answer those cross-cultural questions confronting today's polygraph examiners.

Acknowledgments

This research was sponsored by the Department of Defense Polygraph Institute (DoDPI) project DoDPI89-P-0004, under a contract administered by the Directorate of Contracting, United States Department of the Army (contract DABT02-89-C-0120). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government

References

Anisimov, O. (1951). The Soviet citizen - a profile. *The Russian Review, 10,* 15-23.

Argyle, M. (1982). Inter-cultural communication. In S. Bochner (Ed.). *Cultures in Contact.* New York: Pergamon Press, 75-85.

Barland, G. (1988). The polygraph in practice. In Anthony Gale (Ed.) *The Polygraph Test: Lies, Truth and Science* (pp. 73-95). Beverl7y Hills: Sage Publications.

Barnouw, F. (1985). *Culture and Personality.* Homewood, IL: The Dorsey Press.

Bauer, A. (1952). *The Development History of The Political Attitudes of individuals Toward The Soviet Regime.* Human Resources Research Institute. Technical Research Report No. 6, Maxwell Air Force Base, Alabama.

Bochner, S. (1982). The social psychology of cross-cultural relations. In S. Bochner (Ed.) *Cultures in Contact.* New York: Pergamon Press.

Brislin, R. (Personal Communication, October 17, 1989).

Brislin, R. W. (1983). Cross-cultural research on psychology. *Annual Review of Psychology,* 34, 363-400.

Chan, J. (1979). *The Facial Expressions of Chinese and Americans.* Unpublished doctoral dissertation, South Eastern University, Louisiana.

Cleveland, H., Mangone, G. J., & Adams, J. G. (1960). *The Overseas Americans,* New York: McGraw-Hill.

Cline, R. J.,& Puhl, C. A. (1984). Gender, culture, and geography: a comparison of seating arrangements in the United States and Taiwan. *International Journal of Intercultural Relations, 8*(2), 199-219.

Danton, G. H. (1938). *The Chinese People: New Problems and old Backgrounds.* Boston: Marshall Hones Co.

De Vos, G. (1960). Achievement and innovation in culture and personality. *The Study of Personality: An Interdisciplinary Appraisal.* New York: Holt-Rinehart, 349-370.

De Vos, G. (1968). The relation of guilt toward parents to achievement and arranged marriage among the Japanese. *Psychiatry, 23*(3).

Dicks, H. V. (1952). Observations on contemporary Russian behavior. *Human Relations, 5,* 2. Ann Arbor: The Research Center of Group Dynamics.

Dickson, V. (1949). *Artistic House-Decoration In Riyadh* (Vol. 49). London: The Royal Anthropological Institute.

Drgyle, M. (1982). Inter-cultural communition. In S. Bochner (Ed.), *Cultures In Contact* (pp. 75-85). New York: Pergamon Press.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of emotion. *Motivation and Emotion, 10*(2), 159-168.

Ekman, P., et al. (1987). Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*(4), 712-717.

Ekman, P., Friesen, W. V., & Ellsworth P. (1972). *Emotion in the Human Race.* New York: Pergammon.

Eysenck, H. J. (1982). The biological basis of cross-cultural differences in personality: blood group antigens. *Psychological Reports, 51,* 531-540.

Feldman, R. S. (1983). Nonverbal disclosure of deception in urban Koreans. *Journal of Cross-Cultural Psychology, 10(l),* 73-83.

Gabdreyeva, G., & Peisakhov, N. (1982). A method of self-control over mental states. *Voprosy Psikhologii, 5,* 101-108.

Gorkin, M. (1986). Countertransference in cross-cultural psychotherapy: The example of Jewish therapist and Arab patient. *Psychiatry, 49,* 69-79.

Gudykunst, W. B., & Nishida, T. (1986). Attributional confidence in low- and high-context cultures. *Human Communication Research, 12*(4), 525-549.

Guthrie, G. M., Verstraete, A., Dunes, M. M., & Stern, R. M. (1975). Symptoms of stress in four societies. *Journal of Social Psychology, 95,* 165-172.

Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values.* Beverly Hills, CA: Sage Publications.

Howard, A., Shudo, K., & Umeshema, M. (1983). Motivation and values among Japanese and American managers. *Personnel Psychology, 36,* 355-366.

Johnson, W. R. (1987). The ambivalent polygraph. *International Journal of Intelligence and Counterintelligence, 1*(3).

Kaganiec, H. J. (1956). Lie-detector tests and "freedom of the will" in Germany. *Northwestern University Law Review, 51,* 446-460.

Krober, A. L., & Kluckhohn, C. (1952). *Culture: A Critical Review of Concepts and Definition.* Cambridge, Mass.: Peabody Museum, 47, 1.

Kugelmass, S., & Lieblich, 1. (1968). Relation between ethnic origin and GSR reactivity in psychophysiological detection. *Journal of Applied Psychology, 52*(2), 158-162.

Lazarus, R. S., Tomita, M., Opton, Z., & Kodama, M. (1966). A cross-cultural study of stress-reaction patterns in Japan. *Journal of Personality and Social Psychology, 4,* 622-633.

Lee, D. D. (1953). Greece. In Mead, M. (Ed.), *Cultural Patterns and Technical Change.* Paris: UNESCO.

Levinson, D. (1989) *The human relations area files.* RSR Fall 83-86.

Lieblich, R., Kugelmass, & Ben-Shakhar, G. (1973). Psychochysiological baselines as a function of race and ethnic origin. *Psychophysiology, 10*(4), 426-429.

Linton, R. (1945). *The Cultural Background of Personality.* New York: Appleton-Century.

Malcolm, G. A. (1951). *First Malayan Republic.* Boston: Christopher Publishing.

Mead, M. (1951). *Soviet Attitudes Toward Authority: A Interdisciplinary Approach to Problems of Soviet Character.* New York: McGraw-Hill Book Co., Inc.

Meyjes, P. (1961). Scientific criminal investigation techniques under Dutch law. *Journal of Criminal Law and Police Science, 51,* 643-660.

Montague, A. (1971). *Touching.* New York: Harper and Row.

Ohnishi, K., Matsuno, K., Arasuna, M. & Suzuki, A. (1976). The objective analysis of physiological indices in the field of detection of deception. *Reports of the National Institute of Police Science, 29*(3), 181-188.

Patai, R. (1973). *The Arab Mind.* New York: Charles Scribner's Sons.

Russell, J. A., Lewicka, M., & Niit. (1989). A cross-cultural study of a circimplex model of affect. *Journal of Personality and Social Psychology, 57*(5), 848-856.

Sanover, L. A., Porter, R. E., & Jain, N. C. (1985). *Understanding Intercultural Communication.* Belmont, CA: Wadsworth Publishing Company.

Sharma, S., Parnean, S., & Speilberger, C. (1983). A cross-cultural study of test anxiety levels in Iranian and Indian students. *Personality and Individual Differences, 4*(1), 117-120.

Triandis, H. C. (1980). Introduction to handbook of cross-cultural psychology. In Triandis, H. and Lambert, W. (Eds.) *Handbook of Cross-Cultural Psychology Perspectives.* Boston, Mass.: Allyn and Bacon, Inc.

Triandis, H. (1985). Some major dimensions of cultural variation in client populations. In Pederson, P. (Ed.), *Handbook of Cross-Cultural Counseling and Therapy* (pp. 21-44). Westport, Conn.: Greewood Press,

Triandis, H. C., & Brislin, R. (1988). Cross-cultural training across the individualism - collectivism divide. *International Journal of Intercultural Relations, 12,* 269-289.

Tylor, E. B. (1877). *Primitive Culture. Reseaches into the Development of Mythology, Philosophy, Religion, Language, Art, and Customs.* New York: Henry Holt.

Waid, W. M., & Orne, M. T. (1981). Cognitive, social, and personality processes in the psychophsiological detection of deception. In Berkowitz, L. (Ed.), *Advances in Experimental Social Psychology.* New York: Academic Press.

Wallbott, H. G., & Sherer, K. R. (1986). How universal and specific is emotional experience? Evidence from 27 countries on five continents. *Social Science Information, 25,* 4, 763-795.

Watson, O. M. (1970). *Proxemic Behavior: A Cross-Cultural Study.* Morton, The Hague.

Watson, O. N., & Graves, T. D. (1966). Quantitative research in proxemic behavior. *American Anthropologist, 68,* 971-985.

Wilson, A. T. (1932). *Persia.* London: Ernest Benn Limited.

Wolfgang, A., & Cohen, M. (1988). Sensitivity of Canadian, Latin Americans, Ethiopians, and Israelis to interracial facial expressions of emotions. *International Journal of Intercultural Relations, 12,* 139-151.

Wu, J. C. H. (1967), Chinese legal and political philosophy. In Charles A. Moore (Ed.) *The Chinese Mind,* (pp. 221-235). Honolulu: University of Hawaii Press.

Appendix A
Databases Searched By Computer

EMBASE
(Formerly EXCERPTA MEDICA)

The Excerpta Medica database, EMBASE, is one of the leading sources for searching the biomedical literature. It consists of abstracts and citations of articles from over 4,000 biomedical journals published throughout the world. It covers the entire field of human medicine and related disciplines. The online file corresponds to the 46 specialty abstract journals and 2 literature indexes which make up the printer Excerpta Medica, plus an additional 100,000 records annually that do not appear in printed journals.

LEGAL RESOURCE INDEX

LEGAL RESOURCE INDEX provides cover-to-cover indexing of over 750 key law journals and six law newspapers plus legal monographs. The LEGAL RESOURCE INDEX comprehensively indexes articles, book reviews, case notes, president's pages, columns, letters to the editor, obituaries, transcripts, biographical pieces, and editorials providing access to valuable secondary information for the legal professional and others.

MEDLINE

MEDLINE (MEDLARS onLINE), produced by the U.S. National Library of Medicine, is one of the major sources for biomedical literature. MEDLINE corresponds to three printed indexes: *Index Medicus, Index to Dental Literature,* and *International Nursing Index.* MEDLINE covers virtually every subject in the broad field of biomedicine. MEDLINE indexes articles from over 3,000 international journals published in the United States and 70 other countries. Citations to chapters or articles from selected monographs were also included from May 1976 through 1981.

MEDLINE is indexed using NLM's controlled vocabulary MeSH (medical Subject Headings). Over 40% of records added since 1975 contain author abstracts taken directly from the published articles. Over 250,000 records are added per year.

MENTAL HEALTH ABSTRACTS

The MENTAL HEALTH ABSTRACTS database cites worldwide information relating to the general topic area of mental health. Sources include 1,200 journals from 41 different countries, in 21 different languages, books, monographs, technical reports, workshop and conference proceedings, and symposia. Also included are Far Eastern literature and non-print media.

NTIS

NTIS is available from DIALOG as an online database and in compact-disc format, DIALOG NTIS OnDisc. The NTIS database consists of government-sponsored research, development, and engineering plus analyses prepared by federal agencies, their contractors, or grantees. It is the means through which unclassified, publicly available, unlimited distribution reports are made available for sale from agencies such as NASA, DDC, DOE, HUD, DOT, Department of Commerce, and some 240 other agencies. In addition, some state and local government agencies now contribute their reports to the database.

Truly multi-disciplinary, this database covers a wide spectrum of subjects including: administration and management, agriculture and food, behavior and society, building, business and economics, chemistry, civil engineering, energy, health planning, library and information science, materials science, medicine and biology, military science, transportation, and much more.

PAIS INTERNATIONAL

PAIS (Public Affairs Information Service) INTERNATIONAL is a bibliographic index to the public policy literature of business, economics, finance, law, international relations, government, political science, and other social sciences. It provides references in English to material published worldwide in any six languages: English, French, German, Italian, Portuguese, and Spanish. Approximately 60 percent of the items indexed were originally published in English. It covers printed material in all formats: periodical articles; books; state, local, federal, and non-U.S. government documents; committee hearings, pamphlets; and the reports of public and private organizations. PAIS provides comprehensive coverage of all issues of public policy relating to social, economic, or political problems, including areas such as taxation, multinational corporations, banking, labor, insurance, crime, health, international relations, international trade and specific industries.

PsycALERT

PsycALERT is the companion file to PsycINFO (File 11). PsycALERT provides full bibliographic information and brief indexing for all material subsequently included in PsycINFO. Coverage includes all of the 1,300 journals and serial publications (with the exception of *Dissertation Abstracts International*) that comprise the coverage of PsycINFO. Items will be available for searching on PsycALERT in as little as one to two weeks after the receipt of the item by the American Psychological Association.

PsycINFO

PsycINFO covers the world's literature in psychology and related disciplines in the behavioral sciences. Over 1,300 journals, technical reports, monographic series, and dissertations are scanned each year to provide coverage of original research, reviews, discussion, theory, conference reports, panel discussions, case studies, and descriptions of apparatus.

ONTAP SCISEARCH

ONTAP SCISEARCH contains approximately 30,000 records from early 1986 in File 433, SCISEARCH. ONTAP SCISEARCH provides a low-cost training opportunity for use of a database dealing in the multidisciplinary literature of science and technology. Offline prints are not available in ONTAP files.

SCISEARCH

SCISEARCH is a multidisciplinary index to the literature of science and technology prepared by the Institute for Scientific Information (ISI). It contains all the records published in *Science Citation Index* (SCI) and additional records from the *Current Contents* series of publications that are not included in the printed version of SCI. SCISEARCH is distinguished by two important and unique characteristics. First, journals indexed are carefully searched resulting in the inclusion of 90 percent of the world's significant scientific and technical literature. Second, citation indexing is provided, which allows retrieval of newly published articles through the subject relationship established by an author's reference to prior articles. SCISEARCH covers every area of the pure and applied sciences.

SOCIOLOGICAL ABSTRACTS

SOCIOLOGICAL ABSTRACTS covers the world's literature in sociology and related disciplines in the social and behavioral sciences. Over 1,200 journals and other serial publications are scanned each year to provide coverage of original research, reviews, discussions, monographic publications, conference reports, panel discussions, and case studies.

SOCIAL SCISEARCH

SOCIAL SCISEARCH is a multidisciplinary database indexing every significant item from the 1,500 most important social sciences journals throughout the world and social sciences articles selected from 3,000 additional journals in the natural, physical, and biomedical sciences. SOCIAL SCISEARCH includes many important monographs as well. SOCIAL SCISEARCH covers every area of the social and behavioral sciences.

Appendix B
Training Individuals to Interact With Collectivists

1. Learn to pay attention to group memberships. The Other's behavior depends on norms of the ingroups that are important in the other's life.

2. Keep a close eye on the attitudes of the Other's ingroup authorities. It is likely that the Other's attitudes and behaviors will reflect them.

3. When the Other's group membership changes there is a high probability that the Other's opinions, attitudes, and even "personality" will change to reflect the different group.

4. Spend some time finding out about the Other's ingroups.

5. Do not use yourself as a yardstick of involvement in activities that involve ingroups. The Other is likely to be much more involved with groups than you are used to seeing in your culture.

6. The Other is more comfortable in vertical than in horizontal relationships.

7. If you want the Other to do something, try to see if the Other's superiors can give a signal that they approve of such behavior.

8. If you want the Other to do something, show how such behavior will promote the Other's ingroups.

9. The Other will be uncomfortable in competitive situations.

10. Emphasize harmony and cooperation, help the Other save face, and avoid confrontation.

11. If you have to criticize, do so very carefully. Keep in mind that you cannot criticize the Other's ideas without criticizing the person. In the Other's culture people generally do not say "No" or criticize. They indicate disapproval in very subtle ways.

12. If the Other comes from east Asia, expect extraordinary and unjustified modesty. If you give presentations, consider beginning in a more modest manner than you would in your own country.

13. The Other is likely to be comfortable in unequal status relationships. Status in the Other's culture is likely to be based on age, sex, family name, place of birth and the like. Your social position in your own culture, insignia, and symbols of status count more in the Other's culture than in your own. Do not be shy about displaying them. You position in your own society should be mentioned, so the Other knows how to relate to you. Furthermore, age is an important attribute of status in the Other's culture. It is likely that even small differences in age (e.g., one day older) will result in more respect for the older person. Collectivists will try to convert all horizontal relationship into vertical relationship.

14. When you meet the Other for the first time expect the social behavior to be more formal than you are used to in your country. The behavior will be polite, correct, but not especially friendly. You may have to be introduced to people by someone you know who is also respected by the Other. You have to establish yourself as an ingroup member, by showing proper concern for the ingroup, before the behavior becomes friendly. For example, visiting ingroup members in the hospital, spending free time with ingroup members, giving gifts, and making sacrifices for the group can help establish you as an ingroup member. Then behavior becomes more genuinely friendly.

15. Let the Other guide you toward intimacy. Be willing to disclose personal information, when asked for, but avoid giving information that makes you too different from the Other. However, avoid discussions about sexuality, or any topic that might dishonor the ingroup. Collectivists tend to present themselves in the best possible light and give socially desirable answers much more than do individualists (Hui, in press).

16. Do not jump to conclusions when the Other makes what appears to be a strange suggestion. Try to "play along" until you get more information.

17. Learn to understand illicit behavior. Remember that societies differ in the extent they force people to act or not act in illicit ways. The Other's culture is more likely to tolerate such behavior than is yours.

* Source: Triandis, H.C. & Breslin, R. (1998), *International Journal of Intercultural Relations, 12,* 269-289.

Fiscal Year 2002 Annual Polygraph Report to Congress Department of Defense Polygraph Program

Executive Summary

During Fiscal Year 2002 (FY02) nine Department of Defense (DoD) agencies maintained operational polygraph programs. DoD employs the polygraph as one of many investigative tools to detect deception and assess the credibility of individuals involved in criminal investigations, counterintelligence cases, foreign intelligence and counterintelligence operations, as a condition for access to certain positions or classified information, and for requests for exculpation.

Approximately 74 percent of DoD polygraph tests are conducted as a condition of access to classified information or positions under the DoD Counterintelligence-Scope Polygraph (CSP) Program. The DoD CSP Program is authorized by Public Law 100-180. The purpose of the CSP Program is to deter and detect activity involving espionage, sabotage, and terrorism.

The DoD Polygraph Institute (DoDPI) provides basic and continuing education for all federal polygraph examiners. In FY02, the Accrediting Council of Independent Colleges and Schools accredited DoDPI. The institute currently offers a program in conjunction with Argosy University, Washington, DC to transfer graduate credit from the DoDPI Program towards a Masters Degree in Forensic Psychology, the Forensic Psychophysiology Track. During FY02, a total of 18 federal polygraph examiners were enrolled in this program and 14 were awarded Masters Degrees.

DoDPI also administers the Quality Assurance Program (QAP), which requires that all federal polygraph programs undergo a biennial inspection to ensure that their program meets or exceeds the standards outlined in the Federal Examiner's Handbook. During FY02, QAP inspected 13 federal polygraph programs, including eight non-DoD programs, all of which are now deemed to be in compliance.

The Research Division at DoDPI is staffed by a team of doctoral level scientists. Significant research in and development of alternative tools to detect deception and assess credibility has been limited because the Congressionally appropriated funds for the DoDPI Research Division have not been increased for over ten years. However, the Research Division has made significant progress through collaborative research projects to develop new, and less intrusive, technologies to detect deception and assess credibility.

Just after the close of FY02, the National Research Council (NRC) released a report on the existing scientific evidence of the validity of the polygraph technique. The report recommended that a significant investment in the research and development of new technologies to detect deception and be used for credibility assessment was required to produce tools for these requirements based on sound scientific principles. The Department concurs with this recommendation, and will pursue funding sources to be dedicated to this research.

Finally, it is important to note that the NRC Report also concluded that the polygraph technique is the best tool currently available to detect deception and assess credibility. The Department will continue to use the polygraph technique as it has in the past, until improved technologies or methodologies are developed as a result of scientific research.

I
DoD Polygraph Statistics

The Department of Defense has used the polygraph technique for almost half a century. It is used as one of the tools presently available to detect deception and assess the credibility of individuals involved in criminal investigations, counterintelligence cases, foreign intelligence and counterintelligence operations, exculpation requests, and as a condition for access to

certain positions or information. The polygraph technique is often successful in developing essential information used to resolve national security issues and criminal investigations. A statutory quota of 5,000 for CSP examinations limits DoD use of the polygraph.

The following table reflects DoD Polygraph Program statistics for FY02.

Type of Exam	Number	Percent of Total
Criminal	2,283	19.7%
Exculpatory	354	3.1%
CSP *	8,512	73.6%
All Others**	417	3.6%
Totals* **	11,566	100%

* See page 3 for further breakdown of CSP examinations.

** Includes examinations conducted in support of personnel security investigations, counterintelligence and intelligence operations, and polygraph assistance to non-DoD federal agencies.

*** Does not include polygraph examinations conducted by the National Security Agency (NSA). A breakout of polygraph examinations conducted by the NSA is contained in a classified table submitted with this report. Polygraph examinations conducted by the National Reconnaissance Office are conducted under the authority of the Director of Central Intelligence and are not reported in this report.

II
Fiscal Year 2002
Counterintelligence-Scope (CSP)

Polygraph Examinations

Section 1121 of the National Defense Authorization Act for Fiscal Years 1988 and 1989 (Public Law 100-180, December 4, 1987; 101 Stat. at 1147) authorizes the Department of Defense to conduct CSP examinations as a condition for access to classified information and certain positions.

The purpose of the CSP Program is to deter and detect espionage, sabotage, and terrorism. The following relevant topics are covered during the CSP examination:

- Involvement in espionage or with a foreign intelligence/ security service.
- Involvement in terrorism.
- Unauthorized foreign con-tacts.
- Deliberate failure to protect classified information.
- Damage of sabotage to government information, clandestine collection, or defense systems.

Public Law 100-180 (P.L. 100-180) authorizes DoD to administer CSP examinations to persons whose duties involve access to information that has been classified at the level of top secret or designated as being within a special access program under section 4.4 of Executive Order 12958. This includes military and civilian personnel of the Department and personnel of defense contractors and consultants. Since 1991, the provisions of Public Law 100-180 have limited the Department to 5,000 CSP examinations per year. However, DoD personnel assigned, detailed, or under contract to the following positions are exempted from this ceiling and may be subject to a CSP examination:

- DoD personnel assigned, detailed or under contract with the Central Intelligence Agency.
- DoD personnel assigned, detailed, under contract, or applying for a

position in the National Security Agency.
- DoD personnel assigned to a space where sensitive cryptographic information is produced, processed, or stored.
- DoD personnel employed by, assigned, detailed, or under contract to an office within the DoD for the collection of specialized national foreign intelligence through reconnaissance programs.

The following table provides a breakdown of CSP examinations conducted by the DoD in accordance with Public Law 100-180.

Authority for CSP Exam	**Number of CSP exams**
Special Access Programs	2,819
DIA Critical Intelligence Positions	1,345
Top Secret	0
Exams for Interim Access to Sensitive Compartmentalized Information	55
Total CSP Exams Conducted Under the Congressional Ceiling	4,219
Exempted Examination*	4,293
DoD CSP Program Totals**	8,512

* Includes detailees to CIA and NSA, assignees to cryptographic information processing spaces, and persons in non-NRO reconnaissance programs.

** Does not include CSP examinations conducted by NSA which is reported in a classified table which is submitted with this report. Also does not include examinations conducted by the National Reconnaissance Office, which are conducted under the authority of the Director of Central Intelligence

CSP Refusals

During FY02, two military members declined to submit to CSP testing required as a condition of access to classified information. DoD policy states that those persons who decline to take the CSP examination are denied access to the classified materials in question, but are retained in their position or transferred to other positions in the organization of equal pay and responsibility. Of the two individuals who refused to submit to the CSP examination, one subsequently retired from military service and the other voluntarily left at the end of his term of enlistment.

Specific CSP Examination Results

The polygraph examination results for the 8,512 individuals tested under the DoD CSP Polygraph Program are as follows:

Three hundred seventy-two individuals required more than two series (a series is defined as the collection of at least two polygraph charts on an examinee). A total of 87 examinations required more than one day to complete.

There were 8,245 individuals who were evaluated as no significant response (non-deceptive) to the relevant questions and provided no substantive information. The remaining 267 individuals were evaluated as displaying significant responses (deceptive) and/or provided substantive information.

Two hundred forty-seven individuals made admissions relevant to the issues on the CSP topic list, and through further testing, all relevant issues were resolved favorably to the individual.

Seventeen individuals made admissions relevant to the issues on the CSP topic list, and continued to be evaluated as deceptive during further testing.

Of the 267 individuals who were evaluated as displaying significant responses or provided substantive information, 247

received a favorable adjudication, 11 are still pending adjudication, nine are pending investigation, and no one received adverse action denying or withholding access.

Expansion of the CSP Program

Public Law 100-180 instituted an annual ceiling on the number of CSP examinations conducted by DoD. This annual ceiling has remained at the same level since 1991. Since that time, the Department has identified additional vulnerabilities and threats to classified information that did not exist over a decade ago. The broad based use of information technology systems, coupled with the development of information sharing capabilities over the internet and through other electronic media, require the updating of DoD information assurance policies and practices to keep pace with this emerging threat. DoD is considering the development and implementation of enhanced security requirements for information technology professionals with root access to DoD information systems. These enhanced security requirements may require a CSP polygraph examination for access to DoD information systems. Based on the number of information technology professionals assigned to DoD, an increase in the CSP ceiling to prior 1991 levels may be requested from Congress.

Significant Information Developed

The following cases reflect significant information developed during DoD CSP examinations covered by this report. Most of the information that was developed relates to the removal of classified materials, including computer media, from secure environments. Unauthorized contacts with foreign nationals is another common category of CSP developed information. After information is developed from a subject, further CSP polygraph testing is conducted to determine the extent of the violation. Final CSP polygraph examination results, along with a summary of developed information, are provided to appropriate security officials for further investigation and final adjudication.

♦♦♦♦♦

During a CSP examination, a DoD civilian employee admitted that he removed confidential documents as a souvenir when he left one of his previous assignments. Investigators subsequently recovered the documents from the employee's residence. The employee favorably completed additional polygraph testing.

♦♦♦♦♦

During a CSP examination, an examinee admitted to the inadvertent unauthorized removal and destruction of a classified document. The examinee successfully completed the polygraph examination after this admission.

♦♦♦♦♦

During a CSP examination, the examinee admitted to the removal of classified materials from secure environments on several occasions, so that he could complete office assignments at home. These violations were substantiated through investigation; however, the investigation did not develop any additional information on security violations. The examinee successfully completed a polygraph examination after the close of the investigation.

♦♦♦♦♦

During a CSP examination, the examinee admitted that he had improperly stored classified materials outside of government control, so that he could complete work assignments at home. He denied that any of the materials had been compromised. The examinee surrendered the classified information to investigators. The examinee then favorably completed his polygraph examination.

♦♦♦♦♦

During a CSP examination, a DoD civilian employee disclosed a previously unreported romantic relationship with a foreign national. The employee denied disclosing classified information to his foreign national girlfriend. He successfully completed the polygraph.

♦♦♦♦♦

During a CSP examination, a DoD employee assigned overseas admitted that he had become romantically involved, and eventually co-habitated with a foreign national. The employee also stated that he had discussed classified information with his foreign national girlfriend. After these admissions, he successfully completed the polygraph examination.

♦♦♦♦♦

During a CSP, a DoD employee admitted that he had developed a close relationship with a foreign national while assigned outside the United States. The foreign national served as the employee's clergyman. On one occasion, the clergyman had asked the employee for personal information about another DoD employee assigned to a different overseas site. After making these admissions, the employee successfully completed the polygraph examination.

♦♦♦♦♦

During a CSP examination, a DoD employee admitted to using a restricted government computer system to determine the classified location of the U.S. Navy ship to which his son was assigned. The employee then shared this restricted information with other family members. After providing this information, the employee successfully completed the polygraph examination.

III
Utility of the Investigative Polygraph

During FY02, investigations conducted by DoD obtained significant information from interviews conducted with the aid of the polygraph. DoD policy mandates that the polygraph technique can only be employed to supplement traditional methods of investigation. In the examples detailed below, traditional methods of investigation had been unsuccessful in resolving the matter under investigation. In these cases, information derived from the polygraph examination proved to be invaluable in successful resolution of those matters.

♦♦♦♦♦

A military member was suspected of involvement in a check-kiting scheme after a worthless check was deposited in his checking account, and the money was withdrawn resulting in a loss to the financial institution. When interviewed by investigators, the suspect stated that he believed the worthless deposit was his reenlistment bonus, which had been electronically deposited into his account. After a deceptive polygraph examination, the suspect admitted knowingly depositing the worthless check, and implicated three additional individuals in the check-kiting scheme.

♦♦♦♦♦

A military member was accused of sexual assault by three female trainees under his command. The subject was interviewed and denied any inappropriate behavior with the victims. After a deceptive polygraph examination, the subject admitted to kissing and inappropriately touching all three women.

♦♦♦♦♦

A polygraph examination was administered to a military member implicated in the death of a co-worker. After a deceptive examination, the suspect confessed to killing the victim, and slicing the victim's wrist in an attempt to make the death look like a suicide.

♦♦♦♦♦

Six military members were suspected of stealing two classified laptop computers from a military installation. Five individuals were eliminated as suspects after non-deceptive polygraph examinations. The sixth individual admitted to stealing the computers after a deceptive polygraph examination. The computers were subsequently recovered from

the suspect's residence. Additional polygraph examinations were conducted and determined that the classified information on the computers had not been compromised.

♦♦♦♦♦

A retired military member working as a DoD contractor was administered a polygraph examination to determine the veracity of critical information he provided relating to the terrorist attacks on September 11, 2001. The examinee was evaluated as deceptive, and admitted that he did not know if the information he provided was factual, nor did he know if the event he reported had actually occurred.

♦♦♦♦♦

A supplier to the DoD shipped counterfeit and gray market copier cartridges instead of new cartridges that had been purchased directly from the brand-name manufacturer. The supplier provided a letter of authenticity to the DoD, certifying that the cartridges were purchased directly from the brand-name manufacturer. The employee of the supplier who provided the letter of authenticity was interviewed by investigators, and agreed to submit to a polygraph examination to confirm that he did not alter or counterfeit the letter. After a deceptive polygraph examination, the employee admitted that he had altered a genuine letter of authenticity using his office computer, and submitted it to the DoD. The employee also implicated the owner of the supply company in the scheme.

♦♦♦♦♦

The skeletal remains of a military member missing since 1990 were found concealed in a chimney, located in Germany. An autopsy of the victim could not determine the cause of death, but foul play was suspected. During the course of the investigation, a former military member who had been stationed in Germany at the time of the incident was identified as a suspect. The suspect was interviewed and denied any involvement in the death. A polygraph examination administered to the suspect was evaluated as deceptive. The suspect then admitted that he and four other soldiers had beaten the victim to death and concealed his body in the chimney.

♦♦♦♦♦

An investigation was initiated after five containers of high explosives were stolen from an off-post construction site and recovered several days later on a military installation. A military member was developed as a suspect in the theft. The suspect denied any involvement in the crime and agreed to undergo a polygraph examination. After a deceptive result, the suspect admitted that he and another service member had stolen the explosives and discussed bombing two separate buildings on the installation.

IV
Training and Qualification Standards for DoD Forensic Psychophysiologists (Polygraph Examiners)

DoD maintains stringent standards for polygraph examiners. The basic curriculum taught at DoDPI is based on forensic psychophysiology, and conceptual, abstract, and applied knowledge that meet the requirements of a master's degree level of study. Candidates selected for assignment as a DoD polygraph examiner must meet the following minimum requirements:

- Be a citizen of the United States.
- Be at least 25 years of age.
- Be a graduate of a four-year college.
- Possess two years of investigative or comparable experience with a federal or other law enforcement agency.
- Be of high moral character and sound emotional temperament as confirmed by a background investigation.

- Successfully complete a pre-selection polygraph examination.

- Successfully complete the basic polygraph examiner course at the DoDPI.

All federal polygraph examiners receive their basic training at DoDPI. In FY02, the Institute trained 84 new polygraph examiners. After completing the basic polygraph training course, DoD examiners must intern under the supervision of a certified polygraph examiner for a period of at least six months. In addition, DoD polygraph examiners are required to complete 80 hours of continuing education every two years. To help meet this requirement, DoDPI offers specialized courses in forensic psychophysiology and related disciplines. In FY02, 780 students attended this specialized training.

During FY02, DoD agencies maintained an average of 155 certified polygraph examiners. Over the preceding eight fiscal years, DoD averaged 160 certified examiners on staff per year.

If DoD institutes a CSP polygraph examination requirement for information technology professionals, a significant increase in the number of certified examiners will be required to handle the increased workload.

V
Forensic Psychophysiology (Polygraph) Research

A team of doctoral level scientists staffs the Research Division at DoDPI. Significant research in and development of alternative tools to detect deception and assess credibility has been limited because Congressionally appropriated funds for the DoDPI Research Division have not been increased for over ten years. However, the Research Division has made significant progress toward developing new and less intrusive technologies to detect deception and assess credibility through the use of collaborative research projects.

The National Research Council (NRC) of the National Academies of Science recently released a report, *The Polygraph and Lie Detection (2002).* In the report, the NRC commended recent efforts by the DoDPI Research Division but clearly stated the need for expanding these efforts in the interest of national security. More specifically, the NRC strongly recommends expansion of collaborative research efforts from both independent and government laboratories. The NRC also concluded that no alternative technique has yet been shown to outperform the polygraph technique. As mandated by Congress, DoDPI continues to increase the number of research partners through collaboration to research alternative technologies.

As the country enters a new era of threat, the need to expand beyond traditional polygraph techniques has never been more mission essential. DoDPI hopes to broaden its focus to include the entire spectrum of credibility assessment. This change in focus will allow for the rapid development of technologies designed to assist the intelligence community, the homeland and national security communities, as well as the military services in the Global War on Terrorism. To this end, DoDPI Research Division strives to unite the efforts of the government and scientific communities to give our polygraph examiner workforce leading edge technology in protecting the Homeland and Americans worldwide.

Through the 2002 Broad Agency Announcement (BAA) and staff solicitation, the Research Division accepted 13 new institutional proposals, and two graduate student dissertation/thesis award proposals. DoDPI granted three institutional awards and one dissertation/thesis award. In addition, two awards were granted from the 2001 BAA funds. A total of five proposals were rejected for lack of scientific merit and five proposals are being considered for future awards if funds become available. Additionally, there are six proposals in the review process for future consideration. There are also ten internal DoDPI projects ongoing.

Research projects during FY02 resulted in 35 scientific papers and reports published and available to the community. This is the fourth consecutive year the DoDPI Research Division increased its production of deliverables to our customers. The progressive

increase in activity is a direct result of the recruitment and partnership development with external laboratories.

Collaborative Research Projects

University of New York at **Stony Brook**

An Examination of Response Parameters of Electrodermal Recording to Standard Stimuli. The objective of this project is to investigate whether equivalent electrodermal responses are obtained to equivalent psychological stimuli presented at different electrodermal tonic levels. The outcome will determine whether resistance or conductance is a more accurate measure during PDD examinations. Reviews of final report completed. Waiting for final revision.

Thoughtform Corporation

Development of Improved Automated Scoring Algorithms: An Application of Advance Waveform Combining and Classification Technology to the Analysis of Polygraph Data in Psychophysiological Detection of Deception. This project intends to develop a novel waveform application for the scoring of Relevant/Irrelevant test format polygraph data. A number of advanced statistical methods have been developed for the analysis of complex electroencephalographic (EEG) and magnetoencephalographic (MEG) signals, which may also be applied to polygraph physiological data. The goal of this project is to use traditional polygraph data and combine the data into a novel waveform that will enhance the accuracy of the PDD examiner decision. This technique has been shown to be highly accurate and reliable for single-trial analysis of EEG and MEG signals and should be adapted to similar datasets collected from polygraph examinations. Reviews of final report completed. Waiting for final revision.

Washington School of Medicine and Boeing

Noncontact Sensing of Emotion and Stress Using Laser Doppler Vibrometry. This project involves the use of emerging technologies to develop methods for deriving simultaneous information from the laser Doppler signal regarding multiple physiological functions including body tremor, respiration, cardiac function, muscle contraction, and sweating. Laser Doppler Vibrometry recording methods do not require the attachment of physical transducers, and could be adapted to multiple-examination settings. Awaiting final report. Extension to 2003 has been approved.

University of South Carolina

Research Assistant Professor Research/Research Training in Cognitive Psychophysiology and Detection of Deception. The University of South Carolina conducted research on brain activity as it relates to the detection of deception. The project uses high-definition EEG/ERP recordings and correlates these findings with current autonomic nervous system recordings during a PDD examination. Final report in review.

Washington University

Vericator: Evaluating the Validity of a Voice-based Measure of Detecting Deception. This study assessed the ability of *Vericator*, a computer-based system that evaluates credibility through speech, to detect smugglers at a mock security checkpoint. A U.S. Customs Inspector questioned participants while *Vericator* assessed their veracity. For some the Inspector followed a script of questions without follow-up (Scripted); for the remainder, follow-up questions were permitted (Field-like). Smuggling base rates were 34% and 35%, respectively. The proportions of smugglers correctly identified at the checkpoint, i.e., sensitivity, were low (.12 and .18, respectively). Proportions of non-smugglers correctly identified, i.e., specificity, were much higher (.80 and .78, respectively). Post-hoc analyses produced widely disparate sensitivities (.38 to .78) and specificities (.00 to .96). Final report pending.

University of South Carolina

Research Assistant Professor Research/Research Training in Cognitive Psychophysiology and Detection of Deception. The University of South Carolina is conducting a continuation of research on brain activity as it relates to the detection of deception. The project uses high-definition EEG/ERP recordings and correlates these findings with current autonomic nervous system recordings during a PDD examination. The components include research in cognitive psychophysiology to: 1) continue to refine the localization of cortical sources of deception and utilize these underlying neuronal sources to evaluate deceptive responses at an individual level, 2)

begin separating the process of deception from the effects of workload and attention, and 3) conduct multiple session research and apply statistical approaches to the data to evaluate possible practice and memory effects. Performance period 1 Jun 02 – 31 May 04.

Washington School of Medicine and Boeing

Noncontact Sensing of Emotion and Stress Using Laser Doppler Vibrometry. This is a two-year continuation of a project investigating laser Doppler vibrometry (LDV) as a method for assessing the physiological signs of stress and emotion. The research is aimed at five goals: 1) Continue to develop scientific basis for the method, and continue to explore the range of physiological phenomena that may be of use for the detection of deception, 2) Complete an ongoing study of the somatic and cardiorespiratory responses associated with exposure to affective pictures, 3) Operationalize forcefully the specific state of fear, using standardized social stress methods, and to assess the associated physiology using LDV measures, 4) Translate to a mock crime scenario the critical LDV measures that most successfully characterize the state of fear, 5) Continue to develop and refine the technical basis of the LDV method by upgrading existing hardware and software capabilities. Performance period 1 Oct 02 – 30 Sep 04.

Oklahoma University

Exploring Content Coding Procedures for Assessing the Truthfulness of Verbal Statements. The intent of the research is to examine the utility of verbal analytic procedures in assessing honesty and deception in simulated and real employment contexts. Specifically, criteria from various techniques for assessing the truthfulness and deceptiveness of verbal statements based on their content, quality and expression will be tested in a series of experimental studies. Ongoing data collection and analysis. Performance period 1 Jun 02 – 31 May 05.

The University of Utah

Human and Computer Decision-Making in the Psychophysiological Detection of Deception. The objectives of this study are to develop algorithms to measure the 23 types of physiological changes considered to be diagnostic of deception, assess the stability of those measures across test questions and charts (reliability), and assess their diagnostic validity. It will identify the criteria used by DoDPI-trained examiners to evaluate polygraph charts. It will also explore methods for combining the criteria in order to maximize the accuracy of polygraph decisions. Performance period 1 Jun 02 – 31 Aug 03.

Dissertation/Thesis Awards

The University of Utah

Growth Curve Analysis of Polygraph Data. The polygraph technique is commonly used during criminal investigations to assess the veracity of suspects, witnesses, and defendants. In a typical polygraph examination, a series of test questions is presented from three to five times with a short break between each repetition of the question list. A process known as habituation can produce a progressive decrease in the magnitude of physiological responses over repeated presentations of the test questions. Depending on differences in the rates of habituation for different types of test questions, it may become more difficult or less difficult over the course of a polygraph examination to determine if the examinee is deceptive or non-deceptive. The study will assess if there are diagnostic patterns of change in physiological measures over the course of a polygraph examination. Pending release of funds.

DoDPI Projects

The Relationship Between Facial Skin Surface Temperature Reactivity and Traditional Polygraph Measures Used in the Psychophysiological Detection of Deception: A Preliminary Investigation. This study was designed to investigate the feasibility of combining traditional polygraph measures including blood volume, respiration, and electrodermal activity with facial skin surface temperature (SST) changes recorded using high definition thermal imaging. Participants were randomly assigned to nondeceptive (n=13) or deceptive (n=12) treatment groups using a mock-crime scenario. The frequencies of accurate determinations made using traditional polygraph measures, SST measures, and a combination of polygraph and SST measures were compared using

binary logistic regression. Highest accuracy was obtained using a combination of polygraph and SST measures, suggesting that recordings of facial SST provide information that may be useful when combined with traditional measures during a polygraph examination. These results are discussed in relation to the orienting response (OR) theory proposed by Sokolov, (1963, 1997).

New Decision Rule Development. This study was focused on increasing psychophysiological detection of deception accuracy for the Zone Comparison Test through the modification of the decision rule process. Two two-stage models for producing decisions following conventional physiological data scoring are proposed. The 3T stage uses total score cutoffs to produce decisions, with totals of -6 or less producing decisions of deception indicated, totals of +6 or greater producing decisions of no deception indicated, and totals between these cutoffs producing a no opinion (NO) decision. The 3S stage evaluates scores assigned to individual question pairs in addition to total cutoffs to produce decisions. The assigned scores are totaled for each of three relevant-comparison question pairs, producing a spot score for each question pair. If the total score is -6 or less or if any spot score is -3 or lower, then a decision of deception indicated is produced. If the total score is +6 or higher and if all three spot scores are +1 or greater, a decision of no deception indicated is produced. If neither of these criteria are met, a decision of NO is rendered. The 3T3S model used the total cutoff rule first and the spot score rule only if a NO decision was produced after the 3T stage. The 3S3T model implemented the stages in reversed order. We examined the accuracy of these two-stage models and compared that accuracy with decision models that used only one of the two stages (3T or 3S), using three laboratory and four field data sets. Results showed that across all data sets, the 3T3S and 3S3T models produced 8.3% more correct decisions, 11.5% fewer NO decisions, and 3.1% more incorrect decisions than the 3T and 3S models. Second revision of final report in progress.

Polygraph Question Series Trend Analysis. This project was undertaken to determine whether a different degree of diagnostic value exhibits itself at different points during the question series process. This project will examine this issue using a number of different approaches. First, using hand-scored data from two laboratory studies, the diagnostic contributions from a number of different levels will be considered. These include the relevant-comparison question pair as a function of channel and as a function of question series, all of which to be assessed using a Multiple Regression approach. The contribution of specific question series and data channels will also be assessed, independent of question pair. Second, accuracy rates will be determined using the first three, middle three, and last three question series collected in a five-question series data set. These different, but overlapping data sets will be used to produce decisions using four different decision models. These include one approach using only total cutoffs (3T), one approach using only spot scores (3S), and two approaches using both decision rules in a two-stage process (3T3S and 3S3T). All of this research is exploratory, and conducted to determine empirically if there is an approach equal to or more effective than that produced using the typical three-to-five chart rule in combination with a two-stage decision rule approach. Preliminary data analysis completed.

An Evaluation of TES Decision Rules. In an effort to increase the sensitivity of the TES to deception, decision rules will be applied to field cases to determine how they affect decisions. Score sheet data have been collected from several Federal polygraph programs, and the data are now being converted to electronic form. Ongoing data input.

Comparison of Three Major Scoring Systems. In the field there are currently three principal ways in which polygraph examiners score relevant questions. One method is to score the relevant question against the stronger of the two adjacent comparison questions. This is the practice being taught at DoDPI. A second method is to always score the relevant question against the comparison question that immediately precedes it. The Utah and the Matte scoring system takes this approach. A third major method uses the Either-Or rule, in which the examiner decides

which comparison question to use based on whether there is a strong reaction to the relevant question. The Either-Or rule is taught by Cleve Backster. All three methods are widely used, but the effects of these different systems have not been carefully investigated. In this project we will ask volunteers who are proficient in one of the three systems to score 360 chart segments using the seven-point scoring system. The segments contain one relevant question bracketed by two probable-lie comparison questions. The charts are from single-issue Zone Comparison field cases of criminal suspects. We will assess how total scores are affected by the scoring method, and calculate cutting scores that render similar accuracies for the systems. Scoring complete.

Modified General Question Test Decision Rule Exploration. Previous studies by Senter, Dollins, and Krapohl (2001) and Senter and Dollins (2001) showed that substantial increases in accuracy could be gained within the Zone Comparison Test format by using three to five question series, relative to the standard government practice of using three question series. The current work focused on boosting accuracy through the modification of the decision rule process with the Modified General Question Test (MGQT). A two stage model for producing decisions after score assignment was proposed, in order to compare performance with the pure spot score rule. The first stage uses total score cutoffs to produce decisions, with totals of –6 or lower producing decisions of deception indicated, totals of + 6 or higher producing decisions of no deception indicated, and totals in between these cutoffs producing a no opinion decision. If a no opinion decision is produced using the first stage, the second stage, using a spot score rule, is enacted. In this stage, the assigned scores are totaled for each of the three relevant-comparison question pairs, producing three spot scores. If the total of these three scores is –6 or lower or if any spot is –3 or lower, then a decision of deception indicated is produced. If the total of these three scores is +6 or higher and if all three spots are +1 or greater, a decision of no deception indicated is produced. If neither of these criteria is met, a decision of no is rendered. This project is a retrospective analysis of archived data conducted to determine the effectiveness of this approach with several diverse data sets. This two-stage approach has shown to produce large increases in the number of correct decisions with nondeceptive participants, both with laboratory and field data with the Zone Comparison Test, and the goal is to determine the impact of these rules on MGQT performance. Data Analysis Stage complete.

The Impact of Averaging Assigned Scores on Polygraph Decision Accuracy. The existing body of research on differences between group versus individual performance has produced mixed results. In certain contexts, the group outperforms the individual, and in others, individual performance is superior to that of the group. The present study explored how group performance compared to individual performance in the blind scoring of polygraph question series, using a Zone Comparison Test format. Results provide preliminary evidence that use of a group or collective decision approach in producing polygraph decisions could enhance the validity of the process. Actual implementation of such a process could be achieved using a simultaneous approach where evaluators review polygraph data at the same time as the examiner conducting the exam. Concurrent real time printing or viewing could facilitate such an approach. The blind evaluators could then transmit their assigned totals that could then be combined to produce a group decision, all in time for the examiner to conduct the posttest in the usual fashion. Given the state of high-speed Internet connectivity, such a system that would enable input from evaluators at considerable distance is certainly a possibility. Early report drafting.

Information Processing Associated with Emotional States' Studies. Thermography is a safe, non-invasive technology that measures skin surface temperature on a real time basis. The following studies were designed to determine whether this technology (a special non-contact thermal camera will be used in all four studies) will be useful as a supplement to traditional polygraph measures including blood pressure changes, electrodermal (sweat) activity using metal plates attached to the fingers, and straps attached to the chest and abdomen that records respiratory activity. Cardiovascular activity will be recorded using

either electrodes attached to the wrist and ankle (Studies one and two) or a standard blood pressure cuff (Studies three and four). Participants in the first study will listen to sounds (white noise, tones, clicks) of varying pitch and loudness while their physiological responses are recorded. Participants in the second study will listen to positive (e.g. "ecstatic") and negative ("mutilate") affect words taken from a standard set of emotion word stimuli. Some of the participants will also be asked to complete a counting (math) task varying in difficulty and speed of response. Participants who complete the counting task will also receive feedback about their performance as the task is being completed. Participants in the third and fourth studies will be randomly assigned to one of two groups. Participants in the first group will commit a pretend crime by stabbing a mannequin with a screwdriver and stealing money from a purse. Participants assigned to the second group will be informed that a pretend crime was committed, but that they are innocent. Next, all participants in studies three and four will undergo a polygraph examination using traditional polygraph as well as thermography measures. These measures will be examined to determine whether the thermography measures increase the effectiveness of a polygraph examination. Participants in the fourth study will also be asked follow-up questions (about the participant's involvement in the mock-crime) by the polygraph examiner. While asking these questions, the examiner will be receiving on-line data from the thermal camera to help direct the line of questioning. Four phases. Phase one complete. Other phases proceeding.

Research Projects in Review for FY 2003

Penn State University

The Development of a Computerized Question-Response System and the Use of Reaction Time to Detect Deception. During Year 1, we will develop a computerized system to both present questions and record responses. The two-fold purpose of the computerized system will be to standardize the procedure and to assess the diagnostic utility of both behavioral and physiological reaction time data for the detection of deception using the control question technique (CQT). During year two, a large-scale experiment will be conducted with male and female subjects and examiners to assess the efficacy of the computerized system and the contribution of RT data to detection of deception.

Thoughtform Corporation

Polygraph Decision Support System for Relevant/Irrelevant Format Examinations. There is a need to investigate the validity, reliability, and accuracy of PDD examinations for the Relevant/Irrelevant (R/I) Test Format. There is also a need to better understand and automate the process of determining whether a subject is being truthful or deceptive during a polygraph examination. The development of improved computer algorithms, which accurately process polygraph waveforms to better determine deception or truthfulness, will benefit the field of forensic psychophysiology and terrorist detection.

Boise State University

Effects of Comparison Question Type and Between Test Stimulation on the Validity of Comparison Question Tests. Comparison Question Tests (CQT) are the most commonly used type of psychophysiological detection of deception test in, law enforcement, forensic practice and in national security screening settings (Honts, Raskin & Kircher, 2002; Raskin & Honts, 2002). Such tests play an important role in the Government's national security and law enforcement programs. However, many aspects of the polygraph testing procedure, as it is used in practice, lack strong empirical validation, and in some cases some aspects of the testing procedures lack any empirical validation. Two aspects of comparison test administration are currently the topic of some controversy in polygraph literature. Those areas of controversy concern the type of comparison question used, and between chart stimulation of questions.

University of Nebraska Kearney

Polygraph personnel security screening with subjects for whom English is their second language: The effects of Examiner Language and Interpreter Presence. Although the use of foreign-born employees is not uncommon in the Department of Defense, cross-cultural research in the use of polygraph has been

limited to polygraph examiners' perceptions of differences (Yankee, 1991). Although American and Japanese polygraph examiners suggest that communication issues between examiners and polygraph subjects may lead to inaccuracies, no empirical research has examined whether or not bilingual participants respond differently to the Test of Espionage and Sabotage (TES), of any other polygraph examination format, as a function of language used. With recent increases in pre- and post-employment testing by U.S. Government agencies, such as DOD, CIA, FBI, and DOE and the increased needs and concerns associated with hiring employees not native to the United States, it becomes even more important to makes sure that the current screening procedures used to test employees are accurate with all subjects.

Georgia Institute of Technology

A Radar-Based Non-Contact Physiological Detection of Deception Sensor. GTRI is proposing an emerging technology program (Topic two) to develop a non-contact PDD sensor based on existing Radar Vital Signs Monitor (RVSM) technology. The RVSM is able to detect cardiac function, respiration, muscle contractions, eye blinks, and possibly an indicator correlative with galvanic skin resistance without contact. The frequency of operation of the sensor allows the radar signal to penetrate nearly all types of clothing. The resulting information could be used alone to detect deception, or fused together with the traditional polygraph to increase sensitivity and reduce false positives. The end goals of the proposed program are: Construction of an RVSM prototype for PDD; Development of the necessary foundational science and experiments to examine RVSM information for PDD; Explore fusion of the RVSM with traditional polygraph.

The University of New Brunswick

The Influence of the Threat of a Polygraph Test on Deception Ability. Overall, the argument presented in this paper is that the polygraph interrogation situation is not just a simple measurement occasion. Advocates and critics of the procedure argue that cognitive and psychological schemas involving motivation and emotion are created and modified by the very fact that a polygraph test is included in an interrogation sequence. The contemplation of a polygraph test may well serve a function as an intensifier of stress that acts directly on the interrogation following the polygraph. This intensification may be felt by all suspects but may affect guilty suspects the most. The present study is designed to look at possible effects on behavior indices of deception and truth telling on suspects contemplating a polygraph test.

DoDPI

An Examination of the Impact of Evaluating Additional Question Series Using Field Data. The proposed study is an attempt to validate the findings of Senter, Dollins, and Krapohl (2001), and Senter and Dollins (2002). These previous studies showed statistically significant increases (13.0% and 10.5%, respectively) in the number of correct decisions produced when three or five question series are used to produce decisions relative to when only three question series are used. One limitation of these previous studies is that data were collected using participants in laboratory mock crime scenarios. Thus, the possibility remains that this increase in the percentage of correct decisions may be limited to data collected in laboratory settings, limiting the utility of the technique in the real world. In the present study, we will attempt to replicate the findings of our previous studies using data collected during actual criminal investigations. Roughly 500 sets of question series will be collected from specific issue field cases conducted by examiners employed by the United States Army Criminal Investigation Division Command. Five question series will be collected from each examinee, instead of the usual three. Following the confirmation of deception or truthfulness for each case where possible, blind evaluators who are blind to examinee veracity will assign scores to each relevant-comparison question pair for each data channel and question series for each case. These assigned scores will be used to produce decisions by calculating different combinations of spot versus absolute scores and the three versus three to five question series contingency rule. These decisions will be used to assess accuracy (for confirmed cases) and utility (for unconfirmed cases) for various data collection approaches. To begin early CY '03.

Contracts

Johns Hopkins University

New Feature Development and Countermeasures Detection Improvement. This project is designed to develop improved methods of evaluating physiological data known as features in polygraph examinations. This research is also designed to improve the accuracy of detecting polygraph countermeasures.

Lafayette Instrument

PolyPlot. Computer program for generating and modifying polygraph charts. Construction of a polygraph simulator suitable for use an instructional and research aid for DoDPI. The system will provide DoDPI instructors with the ability to generate unique, realistic charts with various test formats, including ZCT, MGQT, TES, and R/I. The resulting charts can then be used in a classroom environment to demonstrate characteristics of polygraph tracings that may not be readily accessible using existing field data.

Bibliography of the Writings of Dr. Stanley Abrams

Robert Drdak

Abrams, S., & Simmons, G. (2000). Post-conviction polygraph testing: Then and now. *Polygraph, 29*(1), 63-69.

Abrams, S. (1973). The polygraph in a psychiatric setting. *American Journal of Psychiatry, 130*(1), 94-97.

Abrams, S. (1973). The polygraph in Japan. *Polygraph, 2*(1), 36-41.

Abrams, S. (1973). Polygraph validity and reliability: A review. *Journal of Forensic Sciences, 18*(4), 313-326.

Abrams, S. (1973). Problems in interpreting polygraph examinations of malingerers. *Polygraph, 2*(3), 234-238.

Abrams, S. (1974). The psychological basis of polygraph technique. *Polygraph, 3*(4), 377-380.

Abrams, S. (1974). The validity of the polygraph with schizophrenics. *Polygraph, 3*(3), 328-337.

Abrams, S. (1975). A response to Lykken on the polygraph. *American Psychologist, 30*(6), 709-711.

Abrams, S. (1975). Techniques for the expert witness. *Polygraph, 4*(1), 1-6.

Abrams, S. (1975). The validity of the polygraph technique with children. *Journal of Police Science and Administration, 3*(3), 310-311.

Abrams, S. (1976). The control question: A technique for effective introduction. *Polygraph, 5*(4), 290-292.

Abrams, S. (1976). The physiological bases of polygraph transducers. *Polygraph, 5*(2), 146-155.

Abrams, S. (1977). *A polygraph handbook for attorneys*. Lexington, MA: Lexington Books.

Abrams, S. (1977). Polygraphy today. *The National Journal of Criminal Defense, 3*(1), 85-106.

Abrams, S. (1977). A survey of attitudes on the guilt complex technique. *Polygraph, 6*(2), 123-124.

Abrams, S. (1978). The utilization and effectiveness of the stimulation test. *Polygraph, 7*(3), 178-181.

Abrams, S. (1980). An update: Hypnosis and its relationship to polygraphy. *Polygraph Update, 1*(6), 1.

Abrams, S. (1981). A case of multiple personality. *Polygraph, 10*(3), 212-215.

Abrams, S. (1981). Polygraphy. In E. J. Imwinkelried (Ed.), *Scientific evidence* 2nd ed., (pp. 755-804). New York: Practicing Law Institute.

Abrams, S. (1982). Psychophysiological basis of polygraph examinations. *Trauma, 23*(5), 5-31.

Abrams, S. (1984). The penile plethysmograph: A new transducer used for detection and therapy

with sexual deviation cases. *Polygraph, 13*(2), 198-201.

Abrams, S. (1984). The question of the intent question. *Polygraph, 13*(4), 326-332.

Abrams, S. (1987). The heart rate monitor. *Polygraph Update, 4*(1). 1-2.

Abrams, S. (1988). Benefits of polygraphs for child molesters on probation. *NDAA Bulletin, 7*(3), 3.

Abrams, S. (1989). *The complete polygraph handbook.* Lexington, MA: Lexington Books.

Abrams, S. (1989). Probation polygraph surveillance of child abusers. *The Prosecutor, 22*(3, Winter), 29-35.

Abrams, S. (1990). Polygraph supervision of child sex abuse. *The Journal of Polygraph Science, 25*(1), 3-4.

Abrams, S. (1991). The directed lie control question. *Polygraph, 20*(1), 26-31.

Abrams, S. (1991). Polygraphy: A new beginning. *Polygraph, 20*(3), 204-213.

Abrams, S. (1992). A response to Wygant's critique of the discovery test. *Polygraph, 21*(3), 248-253.

Abrams, S. (1995). False memory syndrome vs. total repression. *Journal of Psychiatry and Law, 23* (2), 283-293.

Abrams, S. (1999). A response to Honts on the issue of the discussion of questions between charts. *Polygraph, 28*(3), 223-228.

Abrams, S. (1999). The role of polygraphy with the professional board. *Polygraph, 28*(4), 300-303.

Abrams, S. (2000). Polygraph validity in the new millennium. *Polygraph, 29*(4), 344-356.

Abrams, S., & Abrams, J. B. (1993). *Polygraph testing of the pedophile.* Portland, OR: Ryan Gwinner Press.

Abrams, S., Barefoot, J. K., Marcy, L. P., Weir, R. J., Zimmerman, C. H., & Zonn, L. M. (1974). J. K. Barefoot (Ed.), *The polygraph story: Dedicated to man's right to verify the truth.* Revised 3rd printing ed., (p. 34). Linthicum Heights, MD: American Polygraph Association.

Abrams, S., & Davidson, M. (1988). Counter-countermeasures in polygraph testing. *Polygraph, 17*(1), 16-20.

Abrams, S., Hoyt, D., & Jewell, C. (1991). The effectiveness of the disclosure test with sex abusers of children. *Polygraph, 20*(3), 197-203.

Abrams, S., Marcy, L. P., Weir, R. J., Zimmerman, C. H., & Zonn, L. M. (1972). APA reviews the polygraph, employees and unions. *Security World, 9*(10), 48-52.

Abrams, S., Marcy, L. P., Weir, R. J., Jr., Zimmerman, C. H., & Zonn, L. M. (1972). On the side of truth: APA discusses the polygraph, examiner and test. Part II. *Security World* (July/August), 50-52, 54, 56, 59-60.

Abrams, S., & Ogard, E. (1986). Polygraph surveillance of probationers. *Polygraph, 15*(3), 174-182.

Abrams, S., & Weinstein, E. (1974). The validity of the polygraph with retardates. *Journal of Police*

Science and Administration, 2(1, March), 11-14.

Ansley, N., & Abrams, S. (1980). *The polygraph profession.* Linthicum Heights, MD: American Polygraph Association.

Weinstein, E., Abrams, S., & Gibbons, D. (1970). The validity of the polygraph with hypnotically induced repression and guilt. *American Journal of Psychiatry, 126*(8), 143-146.