Exploration into the Effect of Race on Polygraph Scores and Decisions

Donald J. Krapohl and William B. Gary, Jr.

Abstract

Arther (1998) has asserted that race can influence the profile of response patterns in polygraph testing, specifically in the cardiovascular recordings. There has been virtually no evidence reported of such an effect in polygraphy. To test Arther's observation, multiple analyses were made of the cardiovascular and other polygraph channels using polygraph charts from confirmed deceptive field cases. Arther's assertion of a difference in response profiles between African-American and Caucasian examinees was not supported by any of the analyses.

The influence of race on physiological responding is an important question in the field of psychophysiological detection of deception (polygraphy). The existing body of psychophysiological literature points to some differences in autonomic responsivity among racial groups (Johnson & Landon, 1965; Juniper & Dykman, 1967; Kugelmass & Lieblitch, 1968; Lazarus, Tomita, Opton, & Kodama, 1966; Lieblitch, Kugelmass, & Ben-Shakhar, 1973; Murphy, Alpert, Walker, & Willey, 1988; Sternbach & Tursky, 1965). The race of the examinee has not been generally considered in polygraphy for the interpretation of the physiological recordings, however. The failure to take race into account has invited criticism from scientific groups who have issued formal reports on polygraphy (Office of Technology Assessment, 1983; National Research Council, 2002).

A common defense offered by polygraphers is that traditional polygraphy, using the Comparison Question Technique (CQT), evaluates responding within individual examinees, rather than making comparisons across examinees (Abrams, 1989; Reid & Inbau, 1977). This within-subject approach purportedly ameliorates any cross-racial differences that may exist. There is tentative support for this argument in the work of Buckley and Senese (1991), and Reed (1993), who found no significant effects on overall decision accuracy tied to the race of the examinee.

Though the little available research evidence suggests that there are no racial effects in polygraph decision accuracy, at least one notable writer in the field of polygraphy has asserted that there are differences in the profile of physiological responding between races. Arther (1998) contends that "[t]he cardio reactions of blacks are generally not only much greater but also more valid and reliable than those of whites." Because Arther's statement is taken from a training document, there were no data or citations provided to substantiate this claim. However, Arther has over 50 years of professional practice and writing in polygraphy, and his observations carry substantial weight in the field. Of more importance, if Arther's statement about race and channel-preference is true, racial information might be used to improve scoring systems or automated algorithms. Weighting of channels could be introduced to take advantage of responses that are more valid with one group over another. It became our interest to test Arther's notion on racial differences in polygraphy using the polygraph case database of the Department of Defense Polygraph Institute (DoDPI).

---

1 This article is one in a series under the heading Best Practices. The authors are with the U.S. Department of Defense Polygraph Institute. The views expressed in this paper are those of the authors, and do not necessarily represent those of the Department of Defense or the US government. Reprint requests should be directed to: Donald Krapohl, DoDPI, 7540 Pickens Ave., Ft. Jackson, SC 29207.
Method 1

The DoDPI polygraph case database contains cases conducted on Axcitom computer polygraphs (Axcitom Systems, Houston, TX) by the U.S. government and several local law enforcement agencies. To standardize the sample for this effort, only those cases conducted with the DoDPI Zone Comparison Technique (ZCT) were used, provided that they were single-issue examinations with three relevant questions. From those cases, only those where both race and gender were recorded were selected. It was discovered that in this group there were insufficient numbers of females and confirmed truthful cases, so they were excluded from the present analysis. This left 19 African-American males and 34 Caucasian males for the samples, and all cases were confirmed deceptive.

The “Kircher features” (respiration line length, electrodermal response amplitude, cardiovascular response amplitude) were measured automatically, using software developed for the U.S. government (Extract, ver 3.1). These features have been found to be the most predictive within polygraph tracings (Kircher & Raskin, 1988). The measurements were exported to a spreadsheet for analysis. Significance for all statistical treatments was set at 0.05.

Results 1

The means of the raw measurements of cardiovascular response to relevant questions for the African-American and Caucasian samples were tested for significant differences. None was found ($z = 0.70, p > .05$). Figure 1 below depicts the average cardiovascular reactions for the two groups, with standard error of the mean (SEM) bars.

![Bar diagram showing average cardiovascular response amplitudes and SEM to relevant questions for deceptive African-American and Caucasian males.](image)
It could be argued that measures of average responses may obscure a characteristic of more importance in polygraphy, that of differential arousal to relevant and comparison questions. Arther's (1998) statement may be taken to mean that deceptive African-Americans respond significantly greater to relevant questions than comparison questions in the cardiograph, whereas deceptive Caucasians in this same channel have less of a difference in responding to these two types of questions. Therefore, the more meaningful indicator of racial differences could be a comparative analysis of relative arousal, rather than absolute arousal.

**Method 2**

To test the possibility that races respond differentially to relevant and comparison questions, it was necessary to use a metric for differential arousal. The Objective Scoring System (OSS) was chosen for this task. The Objective Scoring System uses absolute measurements of the "Kircher features" (EDR amplitude, respiration line length, and blood volume amplitude) to create ratios that provide an index of relative arousal to relevant and comparison questions (Dutton, 2000; Krapohl & McManus, 1999). Those ratios are converted to scores with a 7-position scoring system. The final result, the scores, may manifest the racial dissimilarity predicted by Arther.

**Results 2**

Table 1 below shows the OSS scores for these samples. The differences in scores between the races were small, and not significant: pneumograph ($t_{41} = 0.11, p > .05$); electrodermal ($t_{45} = 0.87, p > .05$); cardiovascular ($t_{47} = 1.29, p > .05$). Decision accuracy, using +/-6 as thresholds, are shown in Table 2.

<table>
<thead>
<tr>
<th>Average OSS Scores</th>
<th>African-American</th>
<th>Caucasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumograph</td>
<td>-5.26</td>
<td>-4.97</td>
</tr>
<tr>
<td>Electrodermal</td>
<td>-19.89</td>
<td>-16.47</td>
</tr>
<tr>
<td>Cardiograph</td>
<td>-9.53</td>
<td>-6.97</td>
</tr>
</tbody>
</table>

Table 2. Decision accuracy using the OSS with +/-6 cutting scores for deceptive African-American and deceptive Caucasian examinees. No significant differences between races.

<table>
<thead>
<tr>
<th></th>
<th>DI</th>
<th>NDI</th>
<th>Inc.</th>
<th>Accuracy w/o Inc</th>
</tr>
</thead>
<tbody>
<tr>
<td>African-American (n=19)</td>
<td>17</td>
<td>0</td>
<td>2</td>
<td>100%</td>
</tr>
<tr>
<td>Caucasian (n=34)</td>
<td>30</td>
<td>1</td>
<td>3</td>
<td>97%</td>
</tr>
</tbody>
</table>
Though the OSS scores of the data found no differences in differential responding between African-American and Caucasian examinees, the possibility remained that racial differences in reactions might still exist in a form not captured by the objective measures used in this project. Human evaluators, using field methods, may be able to process the data in ways that would reveal the asserted racial differences in the test charts.

**Method 3**

An experienced polygraph examiner was tasked with evaluating the same charts with the orthodox field methods of global interpretation and numerical scoring. The evaluator was federally certified, and an instructor with DoDPI.

The test charts were printed in the hard copy form familiar to all field polygraphers. The evaluator analyzed the recordings, and provided two types of data. First, he globally assessed the charts, and based on the liability of the cardiograph channel, made forced-choice decisions that the charts were from either African-American or Caucasian examinees. He also scored the charts using the traditional 3-position scoring system (Capps & Ansley, 1992; Harwell, 2000; Krapohl, 1998; Van Herk, 1992).

### Results 3

Table 3 shows the evaluator’s estimate of the race of the 53 examinees. The evaluator’s decision accuracy for race was 52.8%, which was not better than chance ($z = 0.13$, $p > .05$). Therefore, global evaluation of the cardiograph did not prove to be a valid indicator of examinee race for these data.

<table>
<thead>
<tr>
<th>Actual Race</th>
<th>African-American</th>
<th>Caucasian</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race Decision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>6</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>Caucasian</td>
<td>13</td>
<td>22</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>34</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 4. Average manual scores from 53 cases of deceptive African-American and Caucasian examinees by polygraph channel. No significant differences between races.

<table>
<thead>
<tr>
<th>Race</th>
<th>African-American</th>
<th>Caucasian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pneumograph</td>
<td>-0.32</td>
<td>-1.12</td>
</tr>
<tr>
<td>Electrodermal</td>
<td>-5.21</td>
<td>-5.82</td>
</tr>
<tr>
<td>Cardiograph</td>
<td>-1.47</td>
<td>-1.35</td>
</tr>
</tbody>
</table>

*Polygraph, 2004, 33(4)*
The evaluator’s scores were analyzed next. Table 4 lists the average of the manual scores for each channel. There were no significant differences found between the racial groups for pneumograph scores ($\bar{t}$[37] = 1.39, $p > .05$), electrodermal scores ($\bar{t}$[40] = 0.85, $p > .05$), or the cardiovascular scores ($\bar{t}$[42] = 0.16, $p > .05$).

General Findings
The present analysis found that deceptive African-American and Caucasian males did not respond differently on the polygraph from members of the other race, at least in terms of Kircher features, OSS scores, manual scoring, and global interpretation. Scores from African-American examinees were not more or less predictive than those of Caucasian examinees for any of the three polygraph channels. An experienced polygraph examiner was unable to determine the race of the examinee from the polygraph data. There was no evidence that inclusion of racial information in scoring or algorithmic systems would yield any benefit in decision accuracy, nor do the data hint that such differences exist. In view of these findings, and the lack of contrary evidence elsewhere, it is premature to suggest that African-Americans and Caucasians respond differently on the polygraph.

There are limitations in this study that warrant note. First, deceptive cases were used in this study because nondeceptive cases were not available in sufficient numbers. Therefore, the question regarding the racial differences in nondeceptive cases is not addressed here. It remains possible that Arther’s contention regarding racial differences in the cardiograph may apply to those cases.

Similarly, the lack of available female cases also limits the generalizability of this study. Arther makes no distinction for sex in his contention of racial differences in physiological arousal patterns. It remains possible that females do show Arther’s reported racial differences. Buckley and Senese (1991) found that the cardiovascular channel contributed significantly to differences in decision accuracy for truthful African-American females: they were detected at a lower rate than other groups. The Buckley and Senese finding runs contrary to Arther’s predication inasmuch as the cardiovascular channel lowered decision accuracy instead of increasing it for truthful African-American females. Because the sample was quite small, a mere five, Buckley and Senese suggested the finding may have been anomalous. Whether there are racial differences in polygraph data for females remains unresolved.

It is also important to add a remark about our choice to use only confirmed cases rather than testing unconfirmed cases as well. It may be argued that unconfirmed cases are qualitatively different from confirmed cases, and that racial differences that might exist in the field are lost when only the subset of confirmed cases are selected for analysis. Our decision to use confirmed cases was based on a single factor: our interest in comparing physiological reactions with ground truth. To answer Arther’s speculation that “cardio reactions of blacks are generally not only much greater but also more valid and reliable than those of whites,” it was necessary to use cases where validity could be checked. As such, it confined our choice to using cases where ground truth was known. Future researchers may include unconfirmed cases also, to eliminate the possibility that the putative racial effect was not overlooked because of our selection criteria.

In summary, our data did not support Arther’s assertion of racial differences in cardiovascular responsiveness, at least with deceptive African-American and Caucasian males. Work remains to determine whether a racial effect is to be found among truthful and untruthful females.

References


Exploration into the Effect of Race on Polygraph Scores and Decisions


