Initial Investigation of Selected Hypotheses Regarding the Pneumograph in Polygraph Testing

Donald J. Krapohl[1] and Chad Russell[2]

Abstract

The data channels for the standard polygraph (respiration, electrodermal activity, relative blood pressure, vasomotor response) have been individually validated in scientific studies. While most polygraph channels are recorded with a single set of sensors, modern polygraphs record data from two separate pneumograph sensors. The evidentiary basis for the employment of what may be redundant measures is not well documented. The present project involved evaluating six data samples to seek advantages to the use of two pneumograph channels over the use of one. The evidence did not uncover any added value for recording two pneumograph channels using polygraph scores as the metric. Similarly, a second study to examine gender effects did not find compelling evidence to support the use of dual pneumographs. Implications are discussed.

Introduction

While polygraphers used a single pneumograph sensor for the first several decades of the polygraph field, the standard polygraph has included two separate pneumograph channels since the late 1970s. The genesis of this shift may center on the influence of prominent polygraph examiners rather than strictly scientific evidence. Richard Arther and George Harman, two well respected polygraph school directors, came upon this idea in the early 1950s (Arther, 1970; 1989). In his 1970 article, Arther wrote:

> "Starting in 1953, both George Harman and I have continually recorded separately both the chest and upper stomach breathing patterns. Based upon these approximately 32,000 double-breathing examinations involving at least 80,000 tests, *we are convinced that the breathing MUST BE recorded separately from both the chest and the upper stomach! This is because one out of every three times a lying breathing will occur in just one of these two places and not in the other!* Thus, if only one breathing were being recorded, the odds are only 50-50 that the pneumograph tube was at the right place to record the lie. *This means that every polygraphist using just a "one-lung" polygraph will miss a lying breathing reaction one out of every six persons he examines!"* (Emphasis in the original).

Richard Golden and Fred Hunter made a presentation in 1970 that at least partially supported Arther's assertion. In their project

about 13% of deceptive respiration responses would have been missed if the examiner had recorded only the lower pneumograph channel. As with Arther, their analysis did not include an evaluation of numerical scores. Consequently, the value of their findings to practitioners employing more current practices is unknown. Also, all of these reports predate research identifying which pneumograph patterns are reliably associated with deception. For example, Golden and Hunter (1970) counted deep breaths as indicators of deception, a feature later found to be invalid (Kircher, Kristjansson, Gardner, & Webb, 2005). Descriptions of other features Golden and Hunter counted are missing from their report. Given the incomplete reporting of selected features, combined with the counting of invalid features, the relevance of the Golden and Hunter project to current practices could not be determined.

There is published research of the dual recording of pneumographs that could shed light on the issue as to whether the additional instrumentation is repaid with additional diagnostic information. A few years following the Golden and Hunter (1970) presentation, Slowik, Buckley, Kroeker, and Ash (1973) approached the problem using a quantification system based on measurements of certain respiratory features, such as amplitude and duration. Their findings contradicted those of Golden and Hunter, however. They found no statistical differences between the upper and lower respiration data. They concluded that "…if for some reason, the examiner is unable to obtain both respiratory recordings, either one may be considered statistically similar enough to be used independently."

In his seminal work with respiration line length (RLL), Timm (1982a) used measurements of both pneumographs in a laboratory study involving a mock crime. He tabulated the instances where the shortest RLL occurred at relevant and probable-lie comparison questions for truthful and deceptive cases, and looked at the frequency in each cell. Timm found a significant association between RLL and type of question for the truthful and deceptive cases when measuring thoracic respiration, but not with abdominal respiration. In other words,

Timm's data hinted at a difference in diagnostic value of the two pneumograph sensors. However, our reanalysis of Timm's tabled data did not reveal a significant difference in the proportion of correct RLL responses between the two pneumograph channels ($z=1.31$, ns). This would suggest that, within the limits of his sample, Timm did not find a benefit to recording two pneumographs. Viewed from a different perspective, it might suggest that the use of only the thoracic respiration may be adequate. In a separate study published that year Timm chose to use only one pneumograph channel placed over the subject's thorax (Timm, 1982b).

Though the various reports came to different conclusions, none used conventional manual scoring, a very useful and easily understood metric, to determine whether there were useful differences in the data between the two pneumograph sensors. Global impressions and tracing measurements are interesting, but they do not translate directly into practical guidance to today's field examiners.

The research report of Matte and Reuss (1992) was a step forward in that regard. In their research Matte and Reuss analyzed the scores from field cases that had used Matte's eponymous Quadri-Track Zone Comparison Technique. They not only evaluated the upper and lower pneumograph data by ground truth, but also by gender. They characterized some channels as more "productive," approaching this classification system from two different perspectives. One is more global, where each channel was evaluated on "the basis of the clarity and purity of its tracing, and adequacy of its amplitude." The second method was based on "the sum of the verified scores attained for each tracing" with the channel having the highest score consistent with ground truth considered to be the most "productive."

Matte and Reuss (1992) reportedly found differences between the pneumograph channels. Those differences broke along examinee gender. Briefly, males showed responses in the lower pneumograph, or equally between the two pneumograph channels, whereas females tended to respond

in the upper pneumograph or equally between upper and lower.

Matte's results warrant attention. If confirmed in other research, his conclusions could have implications for refinement of algorithms and manual scoring system. For example, if on average males and females manifest more powerfully diagnostic pneumograph information in different pneumograph channels, weighting channels according to the examinee's gender could improve decision accuracy or reduce inconclusives.

It became our interest to examine the assumption that two pneumographs are preferable to a single pneumograph in terms of scores using archives of previously collected data. In the second phase we evaluated the influence of gender on pneumographs, again by analyzing existing data. The overarching goal was to test two field hypotheses regarding the pneumograph: Two pneumograph recordings are better than one, and there are gender differences in pneumograph scores.

Method

Study 1

Research Question: Generally, are two pneumograph tracings better than one?

Data Sources

We obtained the scores from four published research projects. The first was a sample produced by a federal examiner blind scoring 100 laboratory cases, half of which were programmed deceptive, collected during a study by Kircher and Raskin (1988). The technique was the Utah Probable Lie Test, and all examinees were male. While up to five charts were recorded, only the first three charts were considered.

The second set came from Blackwell (1999) who collected scores from three federal examiners evaluating 65 deceptive and 35 truthful field cases. The examinations had been conducted using what is now known as the Federal Zone Comparison Technique (FZCT, Light, 1999). Gender had not been recorded.

The third sample consisted of automated scores from 300 field cases used in the development of the Objective Scoring System, version 2 (OSS-2, Krapohl, 2002). Half of the cases were confirmed deceptive. All of the cases were conducted using the FZCT (Light, 1999). Coding for examinee gender was absent.

The last sample contained field cases using FZCT, and were part of a study previously published on rank order scoring (Krapohl, Gordon & Lombardi, 2008). The original sample consisted of 50 deceptive and 50 truthful field cases. However, software malfunctions resulted in the loss of one case of each type. The remaining 98 cases underwent automated rank order analysis using a system described by Honts and Driscoll (1987). Examinee gender information was not available.

Data Processing

The scores of the two pneumographs are expected to be highly correlated since both correspond with volumetric changes in the examinee's torso at two locations close to one another. However, differences in scores do occur. Because examiners choose only one score to represent respiration when summing scores from all of the polygraph channels, simple decision rules have been developed for those occasions when a score from one pneumograph differs from that of the second pneumograph.

For 7-position scoring, when the two scores differ but are on opposite sides of zero (e.g., -1, +1), examiners use a score of zero. If the two pneumograph scores are not on opposite sides of zero, the examiner uses the score furthest from zero. For example, if the score of the upper pneumograph were -1 and the lower score were 0, examiners would use the value -1 in their computations. For clarity purposes, this score is called "best" here to distinguish it from the strictly upper and lower pneumograph scores. Individual scores were tallied per case for the upper, lower and best pneumographs.

For rank order scoring, the sum of ranks for the relevant questions was subtracted from the sum of ranks of the

comparison questions, resulting in a single value for each pneumograph. These values were averaged within the groups of deceptive and truthful cases.

Significance testing was conducted to determine whether there were differences between pneumograph scores. Alpha was set at .05.

Results

Refer to Table 1. For the first sample of scores from a laboratory study (Kircher & Raskin, 1988), averages among upper, lower and best pneumograph scores differed no more than 0.3 points within truthful and deceptive conditions, that is, 0.3 points per

case. No average scores were significantly different from the other two within either the truthful or the deceptive conditions. No significant differences were observed, suggesting that the use of the "best" score offered no advantage over using the score from just one pneumograph.

Table 2 is a summary of the findings from the Blackwell (1999) data set. Averages among upper, lower and best pneumograph scores differed no more than 0.2 points within either the truthful or deceptive conditions. Within conditions, no average scores were significantly different from the other two. In other words, none of the three sets of pneumograph scores (upper, lower or best) outperformed the other two.

Table 1. Average total scores assigned by case for one federal blind scorer of 50 truthful and 50 deceptive cases from an archive of laboratory cases (Kircher & Raskin, 1988).

	Upper	Lower	Best
Truthful	0.50	0.28	0.44
Deceptive	0.12	0.02	-0.04

Table 2. Average pneumograph scores by case for three blind scorers collected by Blackwell (1999).

	Upper	Lower	Best
Truthful	1.10	1.10	1.24
Deceptive	-0.85	-0.92	-0.96

As noted in the introduction, Arther asserted that in one out of three times a reaction will occur in one pneumograph and not the other. We randomly chose one set of scores from the Blackwell (1999) study, and looked at the frequency of scores for a second pneumograph when the first was assigned a 0. Table 3 lists those tallies.

As Table 4 illustrates, when a 0 was assigned in one pneumograph, a zero score is assigned in the other pneumograph about three-quarters of the time. However, when a

non-0 score was given in the other pneumograph, it was in the correct direction only about 60% of the time, and wrong the other times. To summarize, an overwhelming majority of 0s in one pneumograph produced corresponding 0s in the other pneumograph. In the minority of these cases when the second pneumograph did produce a non-0 score, it was on the right side of 0 a little more than half the time. These data do not endorse the predictions of Arther (1970) that a reaction appears in one pneumograph and not the other about a third of the time.

Table 3. Second pneumograph scores when one of the pneumographs (upper or lower) received a 0 score. Data from one randomly selected scorer of the Blackwell (1999) ZCT cases.

Score	Truthful Cases (n=35)		Deceptive Cases (n=65)	
	Total	Percent of Total	Total	Percent of Total
-3	0	0.0	0	0.0
-2	0	0.0	1	0.0
-1	28	10.5	56	10.4
0	192	71.9	438	81.1
1	47	17.7	44	8.1
2	0	0.0	1	0.0
3	0	0.0	0	0.0
Total	267		540	

Table 4. Percentage of scores for a second pneumograph when the first pneumograph is assigned a 0 score. Data from one randomly selected scorer of the Blackwell (1999) ZCT cases.

	Average %
0	76.5
Correct Direction	14.1
Incorrect Direction	9.3

The data from these two studies (Blackwell, 1999; Kircher & Raskin, 1988) considered manual scores, one for laboratory data and the other of field data. Neither found any differences between the three types of pneumograph scores. However, automated scoring that relies on precise measurement of diagnostic features could provide a more close-grained evaluation of pneumograph data, and reveal differences that escape human evaluators. We next examined pneumograph scores produced by measurements in the form of respiration line length (Timm, 1982a).

Table 5 displays the average pneumograph scores for 150 truthful and 150 deceptive field cases assessed by OSS-2 (Krapohl, 2002). OSS-2 pneumograph scores are based exclusively on RLL. Averages among upper, lower and best pneumograph scores differed no more than 0.7 points within either the truthful or deceptive conditions. As with the scores from Blackwell (1999) and Kircher and Raskin (1988), statistical testing found average scores for upper, lower and best pneumograph were not significantly different from the other two.

Table 5. Average total scores for upper, lower and best pneumograph for 300 field cases used in the development of OSS-2 (Krapohl, 2002).

	Upper	Lower	Best
Truthful	2.73	2.43	2.75
Deceptive	-4.40	-3.83	-4.45

Table 6 shows the average upper and lower pneumograph scores for truthful and deceptive cases when rank order scoring is used. Averages among upper and lower pneumograph scores differed no more than 0.5 points within truthful or deceptive conditions. Statistical tests failed to find a significant difference between upper and lower respiration scores.

Table 6. Average total rank order scores for upper and lower pneumographs for 98 field cases from Krapohl, Gordon and Lombardi (2008).

	Upper	Lower
Truthful	-1.86	-2.27
Deceptive	-7.45	-7.55

Discussion

None of the four data sets found the best pneumograph scores to be significantly different from the upper or lower pneumograph scores, which were not different between themselves. The finding was consistent in both laboratory and field data, and human and automated analysis. Such results correspond with those of Slowik, Buckley, Kroeker, and Ash (1973) and Timm (1982a), but do not support the assertions of Arther (1970; 1989). The findings of Matte and Reuss (1992), which indicated differences in pneumograph scores according to gender, could not be answered in Study 1 due to the absence of gender information or the use of single genders in the studies. Gender is taken up in Study 2.

Based on a general assumption in the field, the "best" pneumograph scores should improve decision accuracy. The present data, looking across all three data sets, suggests this assumption may be incorrect, or the effect is too small to be detected despite the considerable sample sizes used here. On its face there appears to be no or marginal value to dual pneumograph recordings in terms of polygraph scores.

Despite the converging evidence of the three data sets that dual pneumograph tracings do nothing meaningful to polygraph decision accuracy, it would be premature to revisit existing policies regarding polygraph instrumentation. It remains possible that there are gender differences in pneumograph scores which could have been obscured in Study 1 where gender was not an independent variable. Study 2 explores the influence of gender on pneumograph scores.

Study 2

Research Question: Do males and females produce different pneumograph scores?

Data Sources

Two new samples were used in Study 2. The first was created from cases found in the archive of the NCCA confirmed case database. All Federal ZCT cases for which gender was listed were subjected to OSS-2 (Krapohl, 2002), and the pneumograph scores recorded. As with previous data sets, averages were calculated for truthful and deceptive cases for the upper and lower pneumograph scores. Scores were broken out separately by gender. For deceptive cases there were 31 females and 34 males. Truthful cases consisted of 10 females and 18 males.

The second data set came from the study by Kircher, Packard, Bell and Bernhardt (2001). This sample provided two independent sets of manual scores for a laboratory study using the Utah Probable Lie Test. While up to five charts were collected by the researchers, we limited our sample to only three charts. For our analyses the scores from the two scorers were averaged. However, upper and lower scores were not broken out separately. The scorers listed only one score for the pneumograph. This limited our analysis to comparing male and female pneumograph scores by ground truth. There were 42 cases in each group of deceptive females, deceptive

males, truthful females and truthful males for a total of 168 cases. This study also had data for the directed lie, but given previous evidence of the low diagnostic value of pneumograph scores when directed lies are employed (Kircher, Packard, Bell & Bernhardt, 2001) those data were not included here.

Results

Figure 1 displays the average scores, along with error bars, for the sample of OSS-2 scores from the archived cases. There were no significant differences for gender by ground truth for any of the comparisons.

Figure 1. Average total OSS-2 scores and error bars for field cases of male and female truthful and deceptive examinees for upper and lower pneumographs.
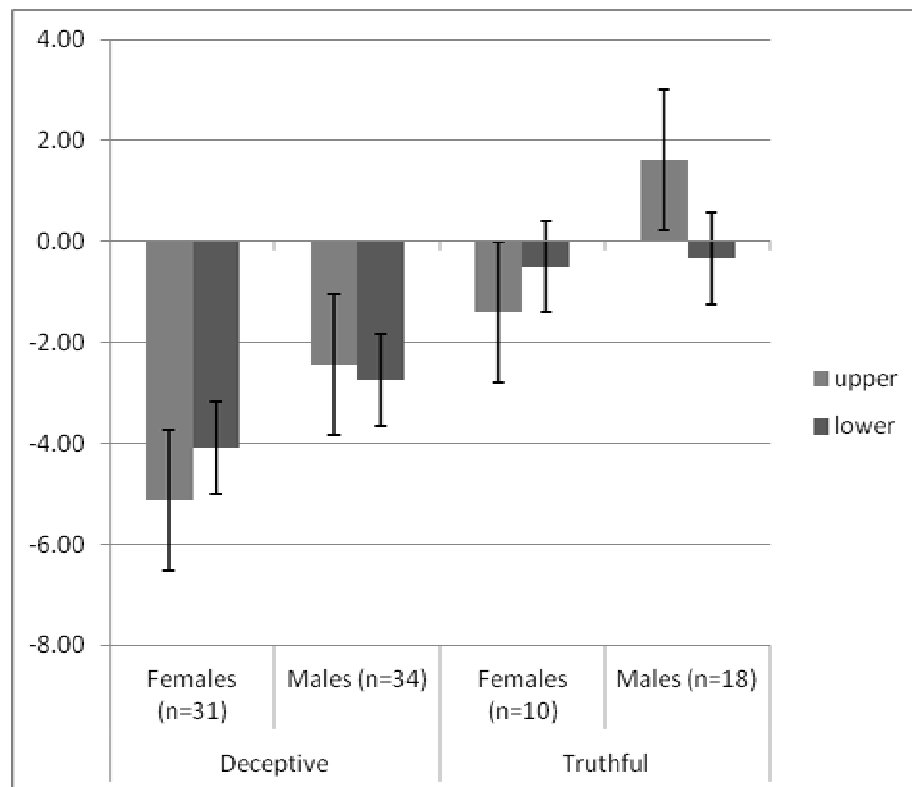
As the error bars indicate, upper and lower pneumograph scores have substantial overlap between males and females. There is a single exception: Truthful female scores in the upper pneumograph are significantly lower than the corresponding scores of males. There is more discussion about this finding under the Discussion section.

Figure 2 compares pneumograph scores for deceptive and truthful females and males. As stated earlier, this data set did not break out upper and lower pneumograph scores separately. The absence of these data allowed us only to consider the overall pneumograph score by gender and ground truth. As the error bars in Figure 2 reveal, there were no significant effects attributable to gender.

65

Figure 2. Average total manual scores and error bars for laboratory cases (Kircher, Packard, Bell & Bernhardt, 2001) of male and female truthful and deceptive examinees.
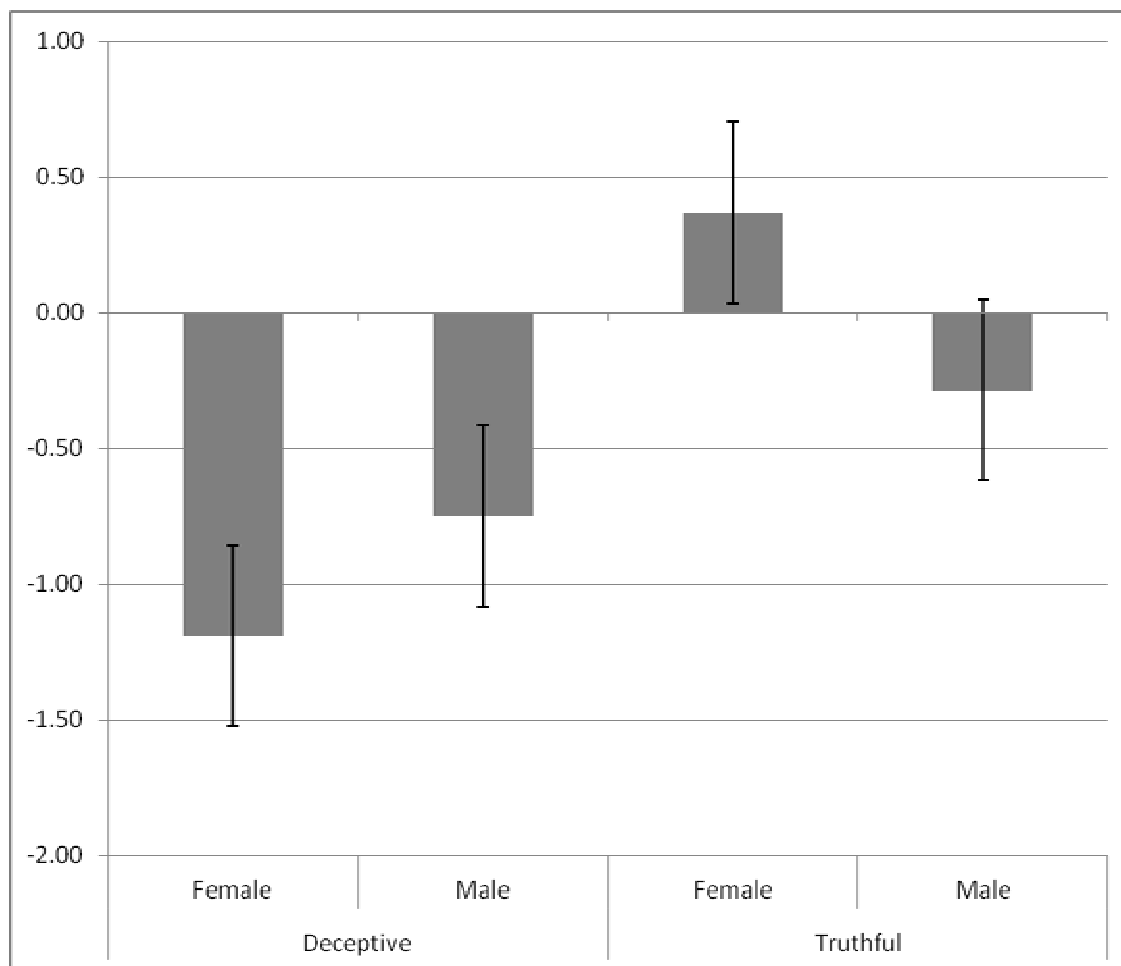
Discussion

Figure 1 hinted to a gender difference between truthful males and females, with the latter producing more negative scores. This is directly opposite the findings of Matte and Reuss (1992) who found females tended to have "more productive" manual scores than males in the upper pneumograph. The conflict between our findings and those of Matte and Reuss (1992) leaves unresolved whether there are differences in upper and lower pneumograph scores associated with gender. It should be noted that given the multiple comparisons in our study, the possibility of a positive finding arising simply from chance alone cannot be dismissed. A similar prospect exists for Matte and Reuss' (1992) opposite finding regarding gender and

pneumographs. Because there is no compelling theoretical reason to predict gender differences in the pneumograph, a conservative interpretation of the conflicting evidence would hold that there is no obvious rationale at this time for using two pneumographs with the expectation that an examiner can extract more diagnostic information based on gender. This view could change if larger and more definitive studies are conducted.

The Kircher, Packard, Bell and Bernhardt (2001) data, though not useful for examination of gender effects on dual pneumographs, failed to find more general effects for gender on single pneumograph scores. Interestingly, a cursory glance at the

Figure 2 histogram bars for truthful males and females shows they go in the opposite direction as those in Figure 1. Though these differences are not significant, the opposite trend lends credence that the statistically significant gender effects for truthful males and females in Figure 1 could be due to normal sample variance.

General Discussion

We found that differences in scores between upper and lower pneumographs appear to be trivial, summed to less than one point per polygraph case for both 7-position and rank order scoring. The insignificant differences also carried over to the average of the "best" pneumograph, the scores of which are normally used in computations to arrive at polygraph decisions. If there is valuable diagnostic value in recording a second pneumograph channel, we failed to find it among all cases or by gender.

It is unclear why other writers have indicated that dual pneumographs provide additional diagnostic information (Arther, 1970; Golden & Hunter, 1970; Matte & Reuss, 1992). It is possible that research that used the more reliable indicators of deception would produce a different tally today. Recall that Golden and Hunter (1970) included deep breaths among the features that they counted as deceptive patterns, something not done today because of advances in understanding in polygraphy. It is also worth noting that their presentation did not include a description of all of the features they were counting.

Regarding the assertions of Arther (1970; 1989) as to the importance of recording two pneumographs, his articles did not include sufficient detail regarding the methodology that led to his conclusions or to permit a replication. Readers are left with his impression, based on his experience with a very large number of examinations, that the absence of the second pneumograph denied examiners of one-third of pneumograph reactions. The only two available research reports that found any effect whatsoever did not report anything approaching this degree of impact for the second pneumograph.

It is possible that Arther's impression of the need for the second pneumograph resulted from the type of instrumentation available in the 1950s when he collected his data, an era before electronic amplifiers were standard on field polygraphs. Amplification now allows examiners to increase signal sizes, and overcome some inadequacies in the quality of the tracings that earlier instruments could not. If instrumentation has provided a remedy to earlier problems with pneumograph signals, the requirement for a redundant system might be obviated.

As early as 1977, Reid and Inbau commented on the recording of dual respiration channels (Reid & Inbau, 1977). In their view, the two channels were "closely parallel to one another" (p. 164), and that reactions in one corresponded with reactions in the other. The value of dual recordings, they claimed, came to detecting when examinees were attempting to distort their responses, a factor Reid and Inbau considered a criterion for a decision of Deception Indicated. In those cases it was the significant departure from the parallelism that Reid and Inbau found meaningful, and such deviations were only detectible when the additional pneumograph channel was recorded (see figures on pages 167 – 168 of Reid & Inbau, 1977, as examples.)

Reid and Inbau also commented on gender differences in the pneumograph tracings in their 1977 text. They wrote:

> Physiologically there appears to be very little difference in respiratory patterns between the sexes. In earlier times, female subjects wore tight corset garments which channeled respiratory responses upward, thus making female subjects appear to be predominantly thoracic responders; at the present time, however, the Polygraph examiner need not make any special allowances when interpreting either abdominal or thoracic response because of a sex difference (p. 167).

As to Matte and Reuss's (1992) more thoroughly described assessment of the two sets of pneumograph scores, we suggest that

there may be a factor that could have influenced their positive finding of sex differences in the scores. As discussed previously, multiple statistical tests of the same data set risks a Type 1 error (false finding of a positive result). In Matte and Reuss' published paper, their analyses entailed 18 separate chi square tests on three variables (upper pneumograph dominance, lower pneumograph dominance, and equal scores) across gender and guilt status. Mere chance can explain some number of positive findings when using such a large number of chi square tests. Similarly, the multiple analyses in the present study could explain the one marginally positive finding in one subset among our six samples.

In addition, we assume that Matte and Reuss based their pneumograph scores on the same features found in Matte's 1996 text. While most of those features have subsequently been shown to be valid, not all have. The use of invalid features for manual scoring is expected to increase the error variance to the extent that the scorer used them, and have an unpredictable effect on subsequent statistical outcomes. It would be informative to replicate the Matte and Reuss (1992) methodology with another data set and use valid scoring features exclusively. This step could resolve questions about their positive pneumograph findings. This same suggestion might also be extended to two of our six samples: the federal scores taken from the Blackwell (1999) and Kircher and Raskin (1988) studies.

Table 7 summarizes the available reports that address the value of dual pneumograph recordings. Despite early endorsements, there remains no unambiguous evidence that would justify the additional instrumentation for what it brings to deception detection. Within the context of conventional scoring we find no practical benefit, nor harm, to the standard practice of recording tandem pneumograph tracings.

Summary

In terms of numerical analysis, our findings showed that each pneumograph channel yields scores highly correlated with those of the other pneumograph channel. The scores assigned to both pneumograph channels individually, for all practical purposes, are redundant measures. Each of the six data sets showed an average difference in scores per case of only a fraction of a single point, making the second pneumograph a negligible contributor to diagnostic accuracy. We also conclude that gender effects in pneumograph scores are, at best, unreliable. From these data we found little basis for employment of the additional instrumentation. It would appear that early claims regarding the value of the second pneumograph channel were overestimated, and those endorsements have little relevance to modern approaches to chart interpretation.

However, as stated by Reid and Inbau (1997), an alternative argument against using a single pneumograph tracing would be the intrinsic value of using dual tracings to assist in identifying intentional distortions of the tracings. Observing the positional relationship of both pneumograph tracings has long been a key identifier in a subject's willingness to cooperate with the test. While parallel or mirrored tracings do not guarantee subjects are offering their full cooperation, they do provide more evaluative data than a single pneumograph tracing could offer. We propose that systematic research could augment field reports in confirming the value of dual pneumographs for this purpose.

Table 7. Summary of research regarding the value of recording a second pneumograph channel.

Researcher	Method of Analysis of the Pneumographs	Supports Dual Pneumographs?	Conclusion
Arther (1970)	Unreported method of tabulating responses in field cases	Yes	One-third of deceptive responses missed with only one pneumograph
Golden & Hunter (1970)	Tallying of deceptive responses (unspecified) in field cases	Yes	13% of deceptive responses missed with only the lower pneumograph.
Slowik, Buckley, Kroeker, & Ash (1973)	Using novel measurement method of responses in field cases	No	No practical benefit to a second pneumograph.
Timm (1982a)	Ranking respiration line lengths from laboratory cases	No	One pneumograph not significantly better than the other.
Matte (1992)	Evaluating manual scores from field cases	Yes	Gender effect: Males most often respond in lower or both, and females in upper or both.
Krapohl & Russell (present article)	Evaluating six separate archived automated and manual scores of field and lab cases.	No	No scoring benefit. Gender effects marginal and not consistent.

References

Arther, R.O. (1970, November-December). The breathing unit. *The Journal of Polygraph Science*, 5(3).

Arther, R.O. (1989, May-June). The breathing unit: An update. *The Journal of Polygraph Science*, 23(6).

Golden, R.I., & Hunter, F.L. (1970). The measurement of upper and lower respiration in lie detection. Paper presented at the American Polygraph Association annual seminar in Los Angeles, CA.

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., Kristjansson, S.D., Gardner, M.K., & Webb, A. (2004). Human and computer decision-making in the psychophysiological detection of deception. Final report to the DoD Polygraph Institute. DTIC ADA520590. Salt Lake City, UT.

Kircher, J.C., Packard, T., Bell, B.G., & Bernhardt, P.C. (2001). Effects of prior demonstrations of polygraph accuracy on outcomes of probable-lie and directed-lie polygraph tests. Report to the DoD Polygraph Institute. DTIC ADA404128. Salt Lake City, UT.

Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology*, 73(2), 291-302.

Krapohl, D.J. (2002). Short Report: Update for the Objective Scoring System. *Polygraph*, 31(3), 298-302.

Krapohl, D.J., Gordon, N., & Lombardi, C. (2008). Accuracy demonstration of the Horizontal Scoring System using field cases conducted with the Federal Zone Comparison Technique. *Polygraph*, 37(4), 263-268.

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

Matte, J.A., & Reuss, R.M. (1992). A study of the relative effectiveness of physiological data in field polygraph examinations. *Polygraph*, 21(1), 1-22.

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

Reid, J.E., & Inbau, F.E. (1977). *Truth and Deception: The Polygraph (Lie-Detector) Technique* (2nd Ed). Williams & Wilkins: Baltimore, MD.

Slowik, S.M., Buckley, J.P., Kroeker, L., & Ash, P. (1973). Abdominal and thoracic pneumograph recordings. *Polygraph*, 2(1), 12-27.

Timm, H.W. (1982a). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10(1), 47-51.

Timm, H.W. (1982b). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *Journal of Applied Psychology*, 67(4), 191-400.