J Forensic Sci, July 2008, Vol. 53, No. 4
doi: 10.1111/j.1556-4029.2008.00775.x
Available online at: www.blackwell-synergy.com

Frank Horvath,[1,†] *Ph.D. and John J. Palmatier,*[2] *Ph.D.*

Effect of Two Types of Control Questions and Two Question Formats on the Outcomes of Polygraph Examinations*

ABSTRACT: Two major variations of polygraph "Control Question" testing, the Zone Comparison (ZoC) and the Modified General Question Test (MGQT) were evaluated. Within each, the type of control question, Exclusive or "time bar" (e.g., "Before you were 21, did you ever...") and Nonexclusive or "no time bar"(e.g., "Did you ever….?") was manipulated in a mock theft scenario, with 80 male and 40 female subjects randomly assigned to be either innocent or guilty. Polygraphic data collected by experienced field examiners were numerically scored by an evaluator blind to all aspects of the study. Decision accuracy was not related to the type of procedure (ZoC/MGQT) used or the subject's sex. Accuracy was significantly related to the type of control question [$\chi^2(2) = 11.46$, $p = 0.003$; $\tau c = 0.29$]. Nonexclusive control questions produced greater accuracy than Exclusive control questions on both innocent and guilty subjects. These results and subjects' self-reports support the general "theory" on which control question (CQ) testing is based. The need for better empirical support of accepted dogma and current field practices is strongly indicated by these findings.

KEYWORDS: forensic science, polygraph testing, control question testing, comparison questions, question formats, detection of deception

Control question (CQ) testing is the most common forensic "lie detection" procedure carried out in field applications in the U.S. and most other countries (F. Horvath, personal communication, July 22, 1996). In this method the truthfulness (innocence) or deception (guilt) of a person on a specified issue is determined by evaluation of physiological response data to relevant and control questions. Simply stated, more pronounced and more consistent physiological responses to control than to relevant questions leads to a decision of truthfulness whereas greater responses to relevant questions leads to a decision of deception.

In the CQ method relevant questions are those directly related to the offense under investigation, e.g., "Did you steal that $500?" Control questions are not directly related to the offense but rather deal with issues concerning the motive for the offense. They are broad in their scope of temporal coverage and they are developed individually with each subject. Their specific form and content is determined by the nature of the interaction between the examiner and the subject. In developing these questions the examiner seeks to interact with the subject in such a way that the subject is led to answer "no" to the question but will have doubt about the truthfulness or accuracy of that answer. In a theft case, for example, a control question might be: "Besides that bicycle you told me about, did you ever steal anything else?" (1). Irrelevant questions, e.g., "Do they call you Joe?" are often used as buffer items to establish baseline response data.

[1]Michigan State University, School of Criminal Justice, 560 Baker Hall, East Lansing, MI 48824.
[2]Dawn Associates LLC, 5732 S Plum Bay Pkwy, Tamarac, FL 33321.
*This research was supported in part by the School of Criminal Justice, Michigan State University and a grant from the American Polygraph Association to the APA's Credibility Assessment Research Center. All statements, opinions, and points of view are those of the authors and do not necessarily represent the position of the APA or its members.
†Present address: 108 Columbia Club Dr. West, Blythewood, SC 29016.
Received 21 May 2007; and in revised form 7 Dec. 2007; accepted 7 Dec. 2007.

The rationale for CQ testing is based on the assumption that persons who are truthful regarding incident-specific (relevant) test questions will be more concerned about the broader nonincident (control) questions, about which they have doubts, than the relevant questions. Deceptive persons, however, will be more concerned about the relevant questions than about the control questions.

In recent years, the CQ testing method has been challenged in the scientific literature (2–4). Some argue, for example, that it is implausible to assume that control and relevant questions will produce differential responses for truthful and deceptive persons. This, however, is an empirical question that can be and has been investigated in the laboratory and in the field (4,5). A recent review of the literature by the National Research Council (NRC) was generally supportive of forensic uses of CQ testing and, equally important, the NRC acknowledged that in spite of its limitations, none of the potential alternatives has yet been shown to outperform polygraph testing (4).

Another argument is that the CQ procedure is not truly standardized; as an example, the "control" question is not a control in the typical scientific sense of that term. Even if true, however, this does not ipso facto invalidate the scientific basis of CQ testing. Nevertheless, recognition of this issue has led to the use of the term "comparison question" instead of "control question" in some literature. In this paper, however, the latter term will be used throughout to be sure that there is a clear connection to the previous literature in the field. (As a sidelight, it is of interest to note that the developer of the control question originally referred to it as a "comparative response question," not a "control" question [6].)

In spite of the controversy about the CQ procedure it continues to be widely practiced and its forensic use is growing (F. Horvath, personal communication, August 23, 2007). It is the predominant method taught in the major training facilities in most countries, including the U.S. Department of Defense's Defense Academy for Credibility Assessment, responsible for the training of all U.S. Federal agency polygraph examiners (F. Horvath, personal communication, August 15, 1995; August 23, 2007). However, as it is

commonly taught, the CQ procedure is not a single, fixed testing method. It is, in fact, a family of conceptually related procedures, generally belonging to one of two major variations, each of which may be preferred depending upon the particular circumstance at hand.

One major variation or format is known as the Zone Comparison (ZoC) method (7,8); the other is the Modified General Question test (MGQT), based roughly on what has been reported by Reid and Inbau (9). Both of these CQ variations are extensively used in the field, although the ZoC format has been, by far, the one most frequently employed in laboratory-based assessments (10). Perhaps, the primary reason for this is that the more recently developed ZoC format is said to incorporate a number of safeguards to deal with habituation and other phenomena (11). In the ZoC method, only three relevant test questions, each of which is merely a rewording of the same question, are typically asked. The three different control questions used are almost always of the Exclusive type (which will be described shortly) and during each repetition of the question list (referred to as a "test"), questions are usually rotated so that each control question is paired with a different relevant question in each test. Evaluation of response data is carried out by scoring only adjacent relevant and control questions (11,12). The ZoC question list also includes a "sacrifice relevant" question, which is not evaluated, hence the name; it is asked to deal with the apparent misleading orienting response said to be elicited by the asking of the first relevant question (1,13). A "symptomatic" question is one asked to detect if a subject's emotional concern is focused on an "outside issue," one not directly related to the offense under investigation (14,15).

The second major CQ variation, the MGQT, may include four or five relevant test questions. Although all are pertinent to the same investigation, each may get at different aspects of it. For example, a subject may be asked: "Did you steal that missing $500 check?", "Did you sign the forged name to that $500 check?", and "Do you know who did steal that $500 check?" It can be seen that the relevant questions in this approach are not necessarily limited to a single issue in the investigation. There are usually only two control questions used and they are often the Nonexclusive type. Because the number of control and relevant questions in the MGQT are not equal, the two types of questions are not immediately adjacent and hence, comparison of responses to relevant questions and control questions is not always of directly adjacent questions. Moreover, a control question response may be compared with more than one relevant question in the same test. Finally, the question list in the MGQT does not normally include the asking of either a sacrifice relevant or a symptomatic question (1,9).

There is a general agreement among field examiners that there should be a rather broad scope of time covered by a control question; this assures that a subject's answer, assuming proper interaction between the examiner and the subject, will have a high probability of being either a lie or, at the least, one which causes the subject concern or doubt about the answer. Beyond this point, however, there are differing schools of thought regarding the nature of the relationship between control and relevant questions. This is the reason for the two types of control questions, Exclusive or Nonexclusive, that predominate in CQ testing. The first school holds that there should be no overlap in coverage between relevant questions and control questions; these Exclusive or time-barred control questions exclude the time period (and sometimes the categorical content, e.g., "Not connected with this case...") covered by the relevant questions. The second type is called a Nonexclusive control question; the scope of coverage of the control question includes the relevant offense. An example of an Exclusive control question (for a 21-year-old person accused of a current theft) would be: "Before you were 18 years of age, did you steal anything?" A Nonexclusive control question in the same situation might be: "Did you ever steal anything in your life?"

The type of control question included in a polygraph examination has been identified as a possible source of error in CQ testing in real-life situations (1,12). In one court case, for instance, the government witness, a presumed authority on polygraph testing, testified that the use of control questions that did not exclude from their scope the time period pertinent to the relevant questions would invalidate the test outcome. Defense witnesses, on the other hand, argued that the use of such questions might reduce the accuracy of the test by a small increment, although they disagreed on the amount of decrease that would occur (16).

In a more recent court hearing, an FBI agent, qualified as an expert, argued against the admissibility of a particular polygraph examination because of the nature of the control questions that were used. He stated that the examination was flawed because it included two control questions that incorporated relevant conduct in their scope of coverage. That, the agent said, was contrary to accepted practice and likely to produce an erroneous outcome (17).

The judge in that case, after hearing experts who differed in their position on the admissibility of the polygraph examination, noted the following:

> …the comparison questions, according to the testing procedure…must be separated by time and date from the issue of the examination. A comparison question should not be structured to include relevant conduct. A comparison question referencing the relevant conduct becomes a relevant question and compromises the examination…This restriction on formatting the questions makes perfect sense because one cannot compare physiological responses between relevant and comparison questions if the comparison question is a relevant question touching on the issue being tested…a properly constructed comparison question would need to avoid any conduct relevant to the [charge at hand]…

It can be seen from this judicial opinion alone that the structure of control questions as they are used in CQ testing is an issue of considerable practical importance. It is also one which is of scientific interest and, fortunately, can be empirically investigated. It is indeed one feature of the CQ method that has been examined in the scientific literature, albeit in a limited way.

The first study to examine this issue was reported by Podlesny and Raskin (12). In their study, a controlled laboratory-based assessment using the ZoC procedure, a direct comparison of the effects on CQ outcomes of two types of control questions was made. The hypothesis was that Exclusive control questions which clearly separate the relevant offense from the scope of the CQ coverage would produce better differentiation of response data and hence, greater accuracy; this is consistent with the generally accepted view of field examiners. Nonexclusive questions, because they incorporate the time period of the offense under investigation, are in reality merely another relevant question, as the judge expressed in the *U.S. v. Williams* decision (17). The use of these types of questions would be more likely to produce either an erroneous outcome (according to the prevailing view, a false-negative error, a deceptive person wrongly reported as truthful, would be most likely the type of error here) or an inconclusive result.

In their findings, Podlesny and Raskin (12) reported that although the difference in the accuracy of the outcomes between Exclusive and Nonexclusive questions was not statistically

significant, the Exclusive type provided some advantages; they, unlike the Nonexclusive control questions led to significant identification of both guilty and innocent subjects and they produced more reliable electrodermal responses. This, it was speculated, was because Exclusive control questions demanded greater information processing wherein subjects had to recall not only whether they had told the truth but also when they had performed the actions specified by the control question.

In a second study reported by Horvath (1), the effects of Exclusive versus Nonexclusive control questions on CQ outcomes were also directly compared. In this study, however, the MGQT method was used. The findings were considerably clearer and in the opposite direction from those specified in the earlier report by Podlesny and Raskin (12). In this study, the Nonexclusive control questions produced more effective identification of both guilty and innocent subjects for each of the polygraphic (physiological) measures, significantly greater confidence in decisions by evaluators on guilty subjects and better control/relevant question differentiation for guilty subjects.

The difference between Horvath's (1) findings and those reported by Podlesny and Raskin (12) is difficult to reconcile. However, one major variable of apparent importance is the CQ testing method (format) that was used in each study. In the Podlesny and Raskin (12) study in which the Exclusive control questions were found to produce better results, the testing method used was the ZoC. On the other hand, the findings of Horvath (1) were produced when an MGQT format was employed. Thus, it might be that the type of control question functions differently depending on the testing format in which it is applied. There has been no research reported in which the MGQT format has been examined in direct comparison to the ZoC method; it was of interest here to do that. In addition, in an effort to clarify the disparity between the Podlesny and Raskin (12) and Horvath (1) studies it was of primary interest to manipulate the control question type. Moreover, because there is a reason to believe that differences in subjects' sex may influence outcomes in certain "lie detection" research, that variable, seldom included in similar studies, was incorporated in the design here (1,4,10,18).

It is commonly recognized that it is difficult to generalize from laboratory findings to the field (1,4,5,12). Yet, there is some recent evidence that suggests this may not be as problematic as some have stated (19). Nevertheless, some variables that may influence the effectiveness of CQ testing can be dealt with much more directly in a laboratory environment, and findings in that context can shed light on testing processes of importance in the field, such as those which might reduce the probability of an error (1,4,20–22). Because it is so widely used and because there is potential for its use in legal proceedings, scientific investigation of the CQ testing procedure and the effects of control question types and format differences on outcomes with that procedure are worthy of attention.

Method

Subjects

Volunteers were recruited from large undergraduate classes in Criminal Justice and Criminology by promising extra credit toward their course work. The students enrolled in these classes were not necessarily majoring in Criminal Justice. They represented areas of study in a wide variety of disciplines of interest to the undergraduate population in the university. The students were told that their participation could result in a small monetary reward if certain conditions of the research, described as involving "lie detection," were met. From the pool of 176 volunteers, 80 white males and 40 white females were randomly selected for participation and each was randomly assigned to one of the predetermined treatment cells in the study, as described below. Males ranged in age from 18 to 27 and females from 18 to 23. The mean age of the males, 19.8 years (SD = 1.65), did not differ significantly from that of females, 19.4 years (SD = 1.22), ($F[1/118] = 1.7$, $p > 0.19$). The remaining volunteers were invited to participate in an alternate extra credit assignment.

Procedure

Each volunteer signed a roster listing his or her name, telephone number, and hours that were available for participation. An assistant contacted each volunteer and provided a date and time for an appointment at a specified office used for administrative tasks related to the research. Upon arrival, an assistant greeted each subject who, after completing an informed consent form and providing other preliminary information, listened to one of two tape-recorded instructional cassettes. Guilty subjects heard a recording that instructed them essentially as follows: "You have been randomly assigned to be a guilty person. Your task, if you choose to participate, is to proceed from this room and to locate Dr. Horvath's office. Once reached, search his mail slot to find a business-size 'airmail' envelope with red and blue markings around the edge and a large red 'X' on each side. When you have this envelope take it out of the slot and hide it; but be careful. Lately there have been thefts of mail from these locations and if someone notices, make whatever excuses that are necessary and continue on your way. You must be careful not to be caught. Return here and you will be given additional instructions and then will undergo a polygraph examination. Do not tell the person conducting the examination what you have done. Any questions related to the envelope, where it was taken from or where you have been must not be answered truthfully. Deny all involvement with the 'airmail' envelope. If you can successfully accomplish this task, that is, if the examination shows that you were telling the truth, you will be rewarded with the contents of the envelope you took. It is important that you do not speak with other students about this study and that you appear truthful at all times. Good luck, now carry out your instructions."

Innocent subjects were greeted, asked to complete an informed consent form, and then they listened to tape-recorded instructions as follows: "You have been randomly assigned to be an innocent person. Your task, once this tape is completed, is to leave the building and return here in 15 min. During the time you are out, a crime will be committed, but you will have no detailed knowledge of what transpired. When you return, you will be subjected to a polygraph examination as a suspect in the crime. You are not to speak with other students about this study and you are to appear as you are, i.e., innocent. If the polygraph examination shows you are being truthful there will be a small cash award in addition to your extra credit. Good luck, now carry out your instructions."

When the guilty subjects returned to the assigned room, they were instructed to display, tear open, and remove the contents of the envelope. In all cases it contained three $1 bills. These were handed to the assistant who then asked the guilty subject to sign his or her name on the "stolen" envelope. When innocent subjects returned to the assigned room they waited there until polygraph testing could be carried out.

Polygraph Testing Procedure

When the examiner was available, each subject was taken to the testing room and introduced by an assistant. The examiner who

was blind to examinees' "guilt" or "innocence," was made aware of which testing protocol was to be used.

One-half of all male (n = 80) and female (n = 40) subjects were randomly assigned to be either "guilty" or "innocent." Within each of these groups, subjects were also randomly assigned to be tested either with the ZoC or the MGQT format and either with Exclusive or Nonexclusive control questions.

Apparatus

As in field settings, polygraph examinations were conducted in a small quiet room (9). Subjects were seated in a commercially available chair with adjustable wooden arm rests. All physiological data were collected with a standard analog field Lafayette polygraph instrument, Model #761-95GA (Lafayette Instrument Company, Lafayette, IN). Two channels of respiration were recorded for each subject by means of two pneumograph tubes. The first was placed around the upper chest (thoracic) and the second around the lower portion of the torso (abdominal). Skin resistance response was recorded in DC mode using stainless steel electrodes attached to the volar surface of the first and second fingers of the left hand and recorded. "Medi-Trace" conductivity gel was used to ensure the best contact possible. Cardiovascular activity was recorded by use of a standard pneumatic blood-pressure cuff, placed around the upper portion of the right arm. Pressure within the cuff was set to 60 mm/Hg and the cuff was squeezed two or three times after inflation to equalize the air within it, generally resulting in a pressure of 40 to 55 mm/Hg. The upper pneumograph and cardio tracings were both enhanced electronically.

Polygraph Examiner

The examiner, a polygraph examiner with a major police agency, had 14 years of police service and 3 years of full-time experience in the administration of polygraph examinations. He had been trained at the Canadian Police College, Canada, a well-known polygraph training facility accredited by the American Polygraph Association. The examiner was trained to use the ZoC test with Exclusive-type control questions. This was also the procedure he used regularly in field settings.

Pretest Interview Phase

A pretest interview similar to that carried out in field settings was conducted with each subject. The examiner offered the reason for the examination—a theft of money from a faculty member's mail slot—and collected relevant biographical and other data such as age, school year, etc. Then, an explanation of the polygraph procedure and the instrument was given. As carried out in the field, control questions were carefully developed with each subject and admissions of conduct within the scope of the question were excluded by appropriate phrasing so that a "no" answer could be given. (In other words, an admission of theft led to a control question such as: "Besides that bicycle [or the items admitted to], did you ever steal anything else?). In all cases, Exclusive and Nonexclusive control questions were prepared in the same way except that when Exclusive control questions were used, the subject was told that the scope was limited to a period of time excluding the 3 years prior to the subject's current age (2). For example, if a subject were 20 years old an Exclusive control question would begin as "Before the age of 17..."

After all control and relevant questions, as required by the assigned treatment condition, were prepared and carefully reviewed with each subject, the polygraph testing was carried out. The order in which tests and questions as well as the number of questions that were presented, varied according to the testing format used, MGQT (1,9) or ZoC (7,8). In either case, a stimulation test commonly practiced in the field was included as part of each examination (1,9).

Administration of the MGQT

The MGQT administered in a manner consistent with its employment in the field, was applied so as to replicate as closely as possible the procedure reported by Horvath (1). In all cases, each "test" included four irrelevant questions, five relevant questions, and two control questions (9). Each relevant question and each control question used in this format, except for modifications necessitated by a subject's admissions and the treatment conditions, was the same for all subjects.

The MGQT question sequence (showing Nonexclusive control questions) was as follows:

1. Do they call you [first name]?
2. Are you over [ ] years of age?
3. Did you take that airmail envelope out of Dr. Horvath's mail slot in Baker Hall?
4. Do you live in the United States?
5. Did you take that envelope containing $3?
6. Did you ever take something that did not belong to you?
7. Did you ever go to school?
8. Did you remove $3 from an airmail envelope taken from Dr. Horvath's mail slot?
9. Did you write your name on that airmail envelope taken from Dr. Horvath's mail slot?
10. Did you ever tell a lie about something important?
11. Were you assigned to be a guilty person in this research?

In this list of questions, questions #1, 2, 4, and 7 were irrelevant questions; questions #3, 5, 8, 9, and 11 were relevant questions; questions #6 and 10 were control questions. All subjects were administered a total of four "tests." The first test was a reading of all questions, in sequence, at 20-second intervals. The examinee answered either "yes" or "no" to each question. The second test conducted was the stimulation test. The third test was a repetition of the first. The final test was a "mixed question" test in which the question order was changed. In this test, the question sequence for all subjects was: #7, 4, 11, 8, 10, 1, 3, 6, 2, 5, 10, and 9. When the testing was finished each subject reported to the assistant at another location to complete a posttest questionnaire to be described shortly.

Administration of the ZoC Test

The ZoC testing format was administered in a manner consistent with its employment in the field. In all instances it included two irrelevant questions, a symptomatic question, a sacrifice relevant question, three relevant questions, and three control questions (19, 32). Excluding the modifications required for the control questions, the question list was the same for all subjects. The question sequence (showing Exclusive control questions, as typically used in this approach) and the question types in the list were as follows:

1. Is your name [first name]?
2. Are you afraid I will ask you a question we have not reviewed?

3. Do you intend to answer truthfully each question about the stolen envelope?
4. Before the age of [ ] did you ever take something that did not belong to you? [Same as MGQT question #6.]
5. Did you take that envelope containing $3? [Same as MGQT question #5.]
6. Before the age of [ ] did you ever tell a lie about something important? [Same as MGQT question #10.]
7. Did you remove $3 from an airmail envelope taken from Dr. Horvath's mail slot? [Same as MGQT Question #8.]
8. Are you now in Michigan?
9. Before the age of [ ] did you ever tell a lie to a person in authority?
10. Did you take that airmail envelope out of Dr. Horvath's mail slot in Baker Hall? [Same as MGQT question #3.]

In this question sequence, questions #1 and 8 were irrelevant questions; question #2 was a symptomatic question; question #3 was a sacrifice relevant question; questions #5, 7, and 10 were relevant questions; and questions #4, 6, and 9 were control questions. As specified in the listing, the three relevant questions were identical to three of the five relevant questions in the MGQT list. Two of the three control questions were the same as the two control questions used in the MGQT listing. All subjects were given a total of four tests. The first test was the stimulation test. The second test consisted of asking of the questions as listed above. The third test was administered with the questions in the following order: #8, 2, 3, 9, 5, 4, 7, 1, 6, and 10. In the fourth test, the question order was again changed to: # 1, 2, 3, 4, 10, 6, 5, 8, 9, and 7. When all testing was completed each subject reported to an assistant at another location.

Evaluation of Polygraphic Response Data

After each subject completed all required tasks the assistant created a file to which a unique code number was assigned. That code number was assigned to each subject's polygraphic charts. After all subjects completed the polygraph testing, the charts were evaluated (numerically scored) by the examiner for the purpose of determining who would receive a monetary award. All subjects whose charts were scored as "truthful" by the examiner were given an award of $3.00, a sum equal to that found in each of the envelopes "stolen" by the "guilty" subjects. Awards were made after polygraph testing of all subjects had been completed.

Because the examiner who carried out the study was aware of the general study design, the particular questions that were asked, and had personally interacted with each subject, blind evaluation of the polygraphic data was carried out by another person, a highly experienced examiner with more than 25 years of full-time work in the field. This evaluator was very familiar with both the ZoC and MGQT formats; in addition, he taught and used field numerical scoring of polygraphic data in his daily activities. He was not a participant in the Podlesny and Raskin (12) or Horvath (1) studies.

All of the subjects' polygraphic charts were sent to the evaluator by mail in groups of ten to twenty, along with predetermined scoring sheets. As a group of charts was scored and returned another was sent to replace it. This process was repeated until all of the polygraph charts were scored.

The only information that was made available to the evaluator was that which could be derived from the polygraphic data. No other knowledge, including the fact that the data were derived from a mock crime (analog) study was made available until all evaluations and all statistical analyses were completed.

The testing examiner, for the purpose of distributing awards, and the blind evaluator recruited for the study, each carried out numerical scoring on each of the 120 subjects' polygraph charts as it would be applied in field settings. This method has been described elsewhere but a brief description is in order (1,12). In this procedure, a relative evaluation of response data for each relevant/control question pair for each physiological measure is made. In considering each pair, the question which is seen to produce a response of greater magnitude (in each measure) determines the sign of a score; a greater response to a control question leads to a positive score whereas a negative score is assigned if the greater response is to the relevant question. Generally, the difference in the magnitude of the relevant questions and control question responses determines the value of the score from 1 to 3 with a large difference leading to a score of 3, a moderate difference a score of 2, and a slight difference a score of 1. If there is no difference, a score of 0 is assigned. Once all pairs have been scored for all physiological measures, the scores are summed across all measures and all of the tests (question list repetitions). In this study, as in many others and in field settings, evaluator total scores of +6 or greater or −6 or less were used as cut-offs to classify subjects as truthful or deceptive, respectively. Total scores falling between ± 5, inclusive, were inconclusive outcomes.

The specific pairs of control and relevant questions that were to be scored were predetermined to ensure that on each chart the same question pairs were considered. These predetermined pairs differed because of the variation in the procedures that were applied, ZoC or MGQT. In the case of the ZoC procedure, each relevant question was compared with the adjacent preceding control question in each of the three tests; thus, because of the rotation of the questions, each relevant question was scored against each control question across the three tests. There were three tests scored, each with three question pairs, each pair being scored for each of four physiological measures; hence, the maximum range possible for the ZoC procedure was between ± 108. MGQT charts were scored using the same relevant/control question pairs applied in the earlier study by Horvath (1). For the first and third tests, these pairs were showing the relevant question number first and the control question number next, as displayed in the earlier listing of the MGQT questions: 3/6, 5/6, 8/6, 9/10, and 10/11. In the last test, a mixed question test, the following pair comparisons were made: 3/6, 5/10, 8/10, 9/10, and 11/10. For the MGQT procedure the maximum range possible was for five question pairs between ± 180 and for three question pairs, of course, the same as in the ZoC procedure. In assigning numerical scores, the evaluator used the following specific criteria, as taught in polygraph training facilities, in assessing response data in the physiological measures.

Respiration—Changes in rhythm or regularity, changes in volume or amplitude, changes in the inhalation-exhalation ratio, notched or serrated respiration strokes, changes in or loss of base line, hyperventilation, suppression, and a holding or blocking (apnea) of respiratory activity.

Skin Resistance—The vertical rise (decrease in resistance) in the tracing, saddle responses, duration of response, and, if present, falling tracing.

Cardiosphygmograph—Increase and decrease in "blood pressure," an increase in blood pressure, a decrease in blood pressure,

increase in pulse rate, decrease in pulse rate, increase in the amplitude of the tracing, decrease in the amplitude of the tracing, a change in the position or presence of the dicrotic notch, and the appearance of extra systoles.

It is to be noted here that both the testing examiner and the blind evaluator "scored," and accumulated in their total scores, the values assigned to each of the two pneumograph tracings. This was carried out because that was the procedure followed by the blind evaluator in his field practice. In addition, such scoring did not permit the blind evaluator to judge arbitrarily which pneumograph tracing yielded the most satisfactory "score" for evaluation, a procedure sometimes applied in field settings.

Concern Ratings

One fundamental premise of the CQ testing is that examinees will be most physiologically aroused by the question or questions that offer the greatest threat to immediate well being. This premise, in the context of CQ testing, provides that guilty persons (relevant question-deceptive) will show greater concern for (be most aroused by) relevant questions because the broader nonspecific control questions offer less immediate threat. Innocent (relevant question-truthful) subjects would show greater concern for the control questions because they are the only questions in the list to which they are either not completely truthful or are in doubt about. To explore the validity of this premise, aside from that which could be shown in the physiological data, each subject completed a short posttest questionnaire in the presence of an assistant, a process also applied in the Horvath (1) study. In this questionnaire, the subject rated on a 4-point scale, from 1 indicating "none" to 4 indicating "a lot," the degree of concern about each of the questions asked during the polygraph testing.

Relevant Questions: Three versus Five

The ZoC procedure employed three relevant questions and three different control questions; the MGQT procedure included five relevant and only two control questions. The effect of this difference in the number of control and relevant questions has not been previously explored, although field examiners, generally, seem to prefer fewer relevant questions (Horvath, F, personal communication, August 15, 1995). In this study, however, three of the five relevant questions employed in the MGQT procedure were identical to the three relevant questions used in the ZoC testing. Thus, it was possible to develop two sets of scores, one which included only the same three relevant questions included in both the ZoC and MGQT procedures and one in which the scores given to the two additional relevant questions in the MGQT procedure were included. Statistical analyses of interest were then possible when the scores were limited to the same three questions included in both procedures and when the scores for the MGQT procedure included the two "extra" relevant questions.

Statistical Analysis

The independent variables of interest were the CQ format employed (ZoC, MGQT); the Type of control question (Exclusive, Nonexclusive); Gender (male, female); and Status (guilty, innocent). The major dependent variables included the numerical scores assigned by the blind evaluator (who was unaware of any details regarding the research design or the subjects), the accuracy of the evaluator's scorings, and the concern ratings assigned by the subjects to the questions they were asked during the testing. In all statistical analyses a 0.05 rejection region was used.

Results

Field Numerical Scoring

To determine the effect of the four independent variables, initial statistical analyses were carried out on the evaluator's total numerical (field) scores. (Correlation of the evaluator's total scores with those of the testing examiner yielded an $r = 0.76$.) It will be recalled that when the ZoC procedure was used, only three relevant questions were included and scored. When the MGQT procedure was used, however, there were two sets of scores that were possible: when only three relevant questions (the same questions used in the ZoC procedure) were scored and when all five relevant questions in the MGQT were scored.

A four-way ANOVA using as factors Procedure (ZoC/MGQT), Status (guilty/innocent), Gender (male/female), and Type of control question (Exclusive/Nonexclusive) was calculated with the total scores of the blind evaluator on the three identical relevant questions used in both formats, serving as the dependent measure. A second calculation was carried out using the scores produced by five relevant questions in the MGQT procedure and three in the ZoC procedure. Neither of these two analyses revealed statistically significant main effects for Procedure or Gender or for any of the interactions involving these factors. However, in both analyses there was a similar pattern of the significant effects. When the same three relevant questions were scored, there was a significant main effect for Status ($F[1/104] = 40.7$, $p = 0.000$) and a significant interaction effect of Status and Type of control question ($F[1/104] = 8.9$, $p = 0.003$). Similarly, when the scores for five questions in the MGQT and the three in the ZoC procedure were subjected to analysis, there was a significant main effect for Status ($F[1/104] = 32.1$, $p = 0.000$) and a significant interaction of Status and Type of control question ($F[1/104] = 5.4$, $p = 0.022$). The mean total scores for both types of control questions are shown separately for the innocent and guilty groups in Table 1. Because neither Procedure nor Gender yielded any significant effects in any of the analyses, the results related to those factors are not displayed and will not be further discussed.

Table 1 shows that whether using the scores of three or five relevant questions (in the MGQT) the mean total scores were more extreme (that is, more strongly in the expected direction) for both the innocent and the guilty subjects when Nonexclusive control questions were used. In other words, in all cases the Exclusive control questions did not discriminate better than did the Nonexclusive questions. It can also be seen in those data that when only three questions were scored, the Nonexclusive control questions were more symmetrical around zero with means of +10.1 and −12.1 than the Exclusive control questions with means of +0.53 and −7.5. A similar finding was found when five relevant questions were scored in the MGQT procedure.

TABLE 1—*Mean total numerical scores (and standard deviations) for innocent and guilty persons using exclusive and nonexclusive control questions.*

	Type of Control Question	
Questions Scored/Status	Exclusive	Nonexclusive
Three questions		
Innocent	+0.5 (15.2)	+10.1 (10.6)
Guilty	−7.5 (12.6)	−12.1 (12.6)
Five questions		
Innocent	+1.9 (20.1)	+11.8 (12.9)
Guilty	−8.4 (14.5)	−12.8 (17.2)

TABLE 2—*Mean total numerical scores (and standard deviations) for each physiological measure for both guilty and innocent subjects.*

	Status	
Measure	Guilty	Innocent
Upper (thoracic) pneumograph	−1.57 (4.7)	0.50 (4.4)
Lower (abdominal) pneumograph	−1.98 (3.8)	0.80 (4.1)
Electrodermal	−4.47 (5.6)	2.60 (5.1)
Cardiovascular	−1.77 (3.3)	1.40 (3.9)

To simplify the presentation of the findings and because the pattern of the results was similar whether the scores for three or five of the relevant questions used in the MGQT procedure were analyzed; only the former results, those for three relevant questions, are hereinafter reported. This, of course, has the additional advantage of ensuring that all of the relevant questions being evaluated were identical in the two procedures, though they were not necessarily located in the same serial position in the question sequence.

Scoring of Physiological Measures

The "total" numerical scores assigned by the evaluator were, of course, an algebraic combination of the scores assigned to each of the four physiological measures. Because there was interest in examining the effects of the treatment manipulations on each of those measures, the "total" scores for each were separately subjected to statistical analysis. These analyses showed that each measure yielded scores that significantly differentiated between the guilty and the innocent; that is, in each analysis, the main effect for Status was statistically significant. In upper pneumograph, lower pneumograph, electrodermal, and cardiovascular analyses, respectively, the main effect for Status (guilty/innocent) was ($F[1/104] = 6.7$, $p = 0.011$); ($F[1/104] = 16.1$, $p = 0.000$); ($F[1/104] = 51.9$, $p = 0.000$); ($F[1/104] = 23.3$, $p = 0.000$). The mean values for each of the scores related to these measures for both guilty and innocent subjects are shown in Table 2.

The main effect for Status revealed in the separate analyses were the only statistically significant findings for the cardiovascular scores; the mean scores shown at the bottom of Table 2 were −1.77 and 1.40 for guilty and innocent subjects, respectively. In other cases, those main effects were confounded by interactions.

For the upper pneumograph, there were two significant interactions, a two-way interaction of Status and Type of control question ($F[1/104] = 8.5$, $p = 0.004$) and a three-way interaction of Status, Gender, and Type of control question ($F[1/104] = 4.62$, $p = 0.034$). The latter outcome is one of only two significant findings in all of the analyses that were related to subject sex. Here examination of mean scores showed that males who were asked Nonexclusive control questions produced only scores in the predicted direction; the mean was 2.20 and −4.00 for innocent and guilty males, respectively. For innocent females, the mean was 1.60; for guilty females, 0.80. All subjects who were asked Exclusive control questions produced negative scores; for innocent male and female subjects, respectively, these mean scores were −1.45 and −0.10. For guilty subjects, the mean scores were −0.55 for males and −1.10 for females.

The scorings of the lower pneumograph measure produced only one interaction effect, a two-way Status by Type of control question outcome ($F[1/104] = 7.6$, $p = 0.007$). Inspection of the mean scores for this effect showed mean values for Exclusive control questions of −0.33 and −1.20 for innocent and guilty subjects, respectively; for Nonexclusive control questions, the mean scores for innocent and guilty subjects were 1.93 and −2.77, respectively. Thus, only the Nonexclusive control questions produced results in the predicted direction.

The scorings of the electrodermal responses yielded only one significant interaction effect, a four-way interaction of Procedure, Status, Gender, and Type of control question ($F[1/104] = 4.47$, $p = 0.037$). This outcome was only the second finding in all of the analyses that was related to the Gender factor and for that reason it will not be further examined here.

Accuracy of Outcomes

The accuracy of the blind evaluator's decisions made by scoring the polygraph charts was of interest. In calculating these statistics, all total numerical scores of −6 or less, or +6 or greater, were used to indicate guilt or innocence, respectively. Total scores between those values indicated an "inconclusive" outcome.

Table 3 shows the number and percentage of correct, wrong, and inconclusive outcomes, collapsed by Status and Type of control question, the only two factors that yielded consistently significant outcomes. Because it was of interest to explore the relationship between the design factors and accuracy, a series of Chi-square tests was carried out on the frequency of correct, wrong, and inconclusive decisions. These analyses revealed that there was a significant moderately strong relationship between accuracy and the Type of control question (χ^2 [2] = 11.46, $n = 120$, $p = 0.003$; $\tau c = 0.29$). The data pertaining to this finding can be seen in Table 3, which also shows the overall percentage accuracy for both innocent and guilty subjects when, as is commonly done, inconclusive outcomes are eliminated. It can be seen in those data that both types of control questions produced a similar accuracy, inconclusives excluded, on guilty subjects, 85% and 80% for Nonexclusive and Exclusive questions, in order; this difference was not statistically significant (χ^2 [1] [$n = 47$] = 0.219).

The difference in accuracy on innocent subjects for the two types of control questions, inconclusive decisions excluded, was statistically significant (χ^2 [1] [$n = 45$] = 11.03, $p = 0.001$, $\tau b = 0.50$).

TABLE 3—*Number and percentage of decisions by status and type of control question.*

	Status							
	Innocent				Guilty			
Type of Control Question	Corr	Wrong	Inc	% Accuracy*	Corr	Wrong	Inc	% Accuracy†
Nonexclusive	21	2	7	91	23	4	3	85
Exclusive	10	12	8	45	16	4	10	80

Corr, correct; Inc, inconclusive.

*Accuracy on innocent subjects, inconclusives excluded, significantly differed by Type of control question (χ^2 [1] [$n = 45$] = 11.03, $p = 0.001$, $\tau b = 0.50$).

†Accuracy on guilty subjects, inconclusives excluded, was not significantly different for the two types of control questions (χ^2 [1] [$n = 47$] = 0.219).

On these subjects, the Nonexclusive control questions produced an accuracy of 91% whereas that statistic for the Exclusive control questions was 45%. The same pattern of statistically significant findings was revealed when inconclusive outcomes were not excluded from the calculations.

Decision Cut-off Scores

Because of the pronounced and statistically significant findings related to the Type of control question on both the accuracy of decisions and the numerical scores, it was of interest to explore the effect of altering cut-off points on outcomes. It will be recalled that scores of +6 or greater and −6 or less were used to determine innocence or guilt, respectively; scores falling between those two points produced inconclusive outcomes.

Using symmetrical and identical cut-off points for both Types of control questions might not have been optimal. To examine that issue, the percent accuracy and inconclusive outcomes was calculated for each type of control question for the guilty and the innocent subjects. These results are shown in Figs. 1 and 2 for the guilty and innocent subjects, respectively. All results are based on the use of the same three relevant questions in both the ZoC and MGQT procedures.

Figure 1 shows the percent accuracy and inconclusive decisions for the guilty subjects. There it can be seen that the Nonexclusive control questions produced a smaller proportion of inconclusive decisions and a higher accuracy than did the Exclusive control questions across the range of cut-off scores from 0 to −8. Further inspection of those data suggests that the optimal (high accuracy, relatively low levels of inconclusive outcomes) cut-off range for

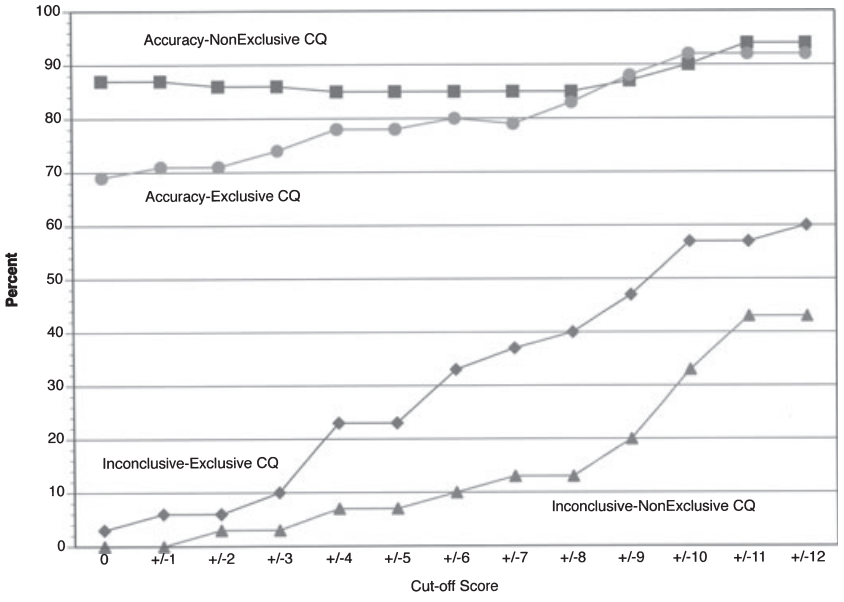

FIG. 1—*Accuracy and inconclusive outcomes for guilty subjects at different cut-off scores.*

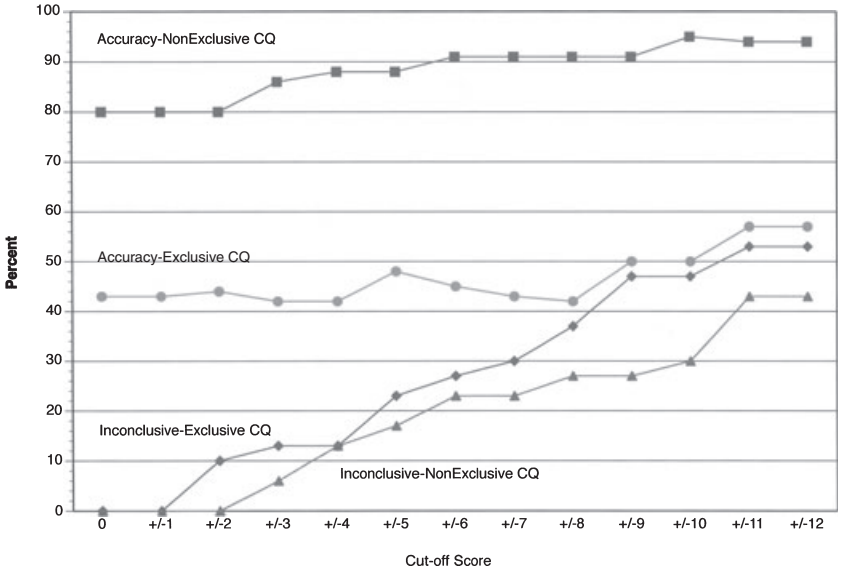

FIG. 2—*Accuracy and inconclusive outcomes for innocent subjects at different cut-off scores.*

the Nonexclusive control questions was between ± 4. For the Exclusive control questions, the preferred cut-off would appear to be similar; higher cut-off scores produced a large number of inconclusive decisions.

Figure 2 showing the effect of cut-off scores between 0 and ± 12 for innocent subjects reveals the clear superiority of the Nonexclusive control questions over the Exclusive control questions across the range of cut-offs. The optimal cut-off score for the Nonexclusive control question was in the range of ± 2 to ± 4, showing that cut-off scores for this question type were relatively symmetrical for both categories of subjects. The optimal cut-off for Exclusive control questions on innocent subjects appears to be between ± 1 and ± 4 though the accuracy in that range was quite low.

To further explore optimal cut-off scores, an index of diagnostic accuracy, a measure of detection efficiency reported by Kircher et al. (10), was calculated separately for Exclusive and Nonexclusive control questions for each cut-off score from 0 to 10. This statistic, a correlation coefficient showing the strength of the relationship between evaluator decisions and the criterion of guilt or innocence revealed that when Nonexclusive control questions were used, detection efficiencies ranged from 0.66 to 0.70 and was highest, 0.70, at cut-off scores of both ± 3 and ± 4. For Exclusive control questions, detection efficiencies ranged between 0.10 and 0.26 and was highest (0.26) at a cut-off score of ± 9. These statistics reinforce what can be seen in Figs. 1 and 2, that optimal cut-off scores differ as a function of the type of control question that was asked.

Subject Concern Ratings

It will be recalled that immediately following their polygraphic examination, each subject indicated the degree of perceived concern during the testing for each relevant and control question on a scale ranging from 1 ("None") to 4 ("A lot"). These ratings were analyzed using a four-way MANOVA, with Procedure (MGQT/ZoC), Status (guilty/innocent), Gender (male/female), and Type of control question (Nonexclusive/Exclusive) as independent variables and the subject's "concern ratings" on relevant questions and control questions as repeated measures ("within subjects"). This analysis revealed only two statistically significant effects, a main (between subjects) effect for Status ($F[1/104] = 34.15$, $p = 0.000$) and a within-subjects interaction effect for Status by question category (Relevant/Control) ($F[1/104] = 98.61$, $p = 0.000$). The mean values (and SDs) for this latter finding are shown in Table 4 where it can be seen that, as suggested by the findings pertaining to the scoring of the physiological data, and by what might be referred to as "CQ theory," the innocent subjects expressed greater concern for the control questions ($M = 2.28$) than for the relevant questions ($M = 1.51$); the guilty subjects rated the two question types in an opposite direction with mean values of 2.89 and 2.06 on the relevant and control questions, in order; calculation of η^2 showed an effect size of 0.48. It is to be noted here that though the "concern ratings" did not significantly differ by Type of control question, in all instances the ratings given for the Nonexclusive control questions were more extreme in the predicted direction than those for the Exclusive questions.

TABLE 4—*Mean subject concern ratings (and standard deviations) by status and type of question.*

	Question Category	
Subject's Status	Control	Relevant*
Innocent	2.28 (0.79)	1.51 (0.73)
Guilty	2.06 (0.68)	2.89 (0.82)

*The interaction of status (Innocent/Guilty) by question category (Relevant/Control) was statistically significant ($F[1/104] = 98.61$, $p = 0.000$; $\eta^2 = 0.488$).

TABLE 5—*Subject's self-reports of attempts to "beat the test."*

	Try to Beat Test	
Subject's Status	Yes	No*
Innocent		
n (%)	2 (3%)	58 (97%)
Mean score (SD)	15.5 (07.8)	4.9 (13.9)
Guilty		
n (%)	24 (40%)	36 (60%)
Mean score (SD)	−10.4 (11.9)	−9.4 (13.3)

*Significantly more guilty subjects than innocent subjects tried to "beat the test", (χ^2 [1] [$n = 120$] = 23.76, $p = 0.000$; $\phi = 0.44$).

Efforts to Defeat the CQ Testing

In addition to indicating their degree of concern for the test questions, each subject also indicated whether or not anything was performed during the testing in an effort to "beat the test." The question to which they responded was to be answered with either a "yes" or a "no," and if "yes" the subject was asked to indicate what was performed, e.g., mental or physical efforts. Even though there were several "Yes" responses, few subjects' gave an explanation that was useful for further analysis. Consequently, only the distribution of "yes" and "no" responses and the numerical scores of the evaluator to these subjects were examined further. As shown in Table 5, less than half of the guilty subjects ($n = 24$) acknowledged trying to "beat the test;" whereas only two of the innocent subjects made a similar acknowledgment. This difference was statistically significant ($\chi^2[1]$ [$n = 120$] = 23.76, $p = 0.000$). Further examination of these subjects' data showed that the mean scores of those who were actually guilty and tried to "beat the test" produced numerical scores that were more "deceptive." In other words, these naïve attempts by the guilty persons actually made them somewhat more, not less, detectable. The number of innocent subjects who said they tried to "beat the test" (presumably these subjects were trying to make certain that they did not produce an erroneous outcome) was too small to determine with certainty the effect of their efforts.

Discussion

The two factors that had the most consistent and largest effects on outcomes were the subjects' Status (guilt/innocence) and the Type of control question (Exclusive/Nonexclusive) that was asked. It is the latter issue, in particular the disparity between the findings of Podlesny and Raskin (12) and those of Horvath (1), as noted at the outset of this paper, that was of primary interest. The results here replicate Horvath's report. Nonexclusive control questions were demonstrably more effective. Contrary to the Podlesny and Raskin (12) report, Exclusive control questions did not yield better results with innocent (truthful) persons; speculation that the use of such questions is to be preferred because they require more information processing was not supported in these data. Moreover, there was no interaction between the Type of control question and the testing formats, ZoC or MGQT, that were employed in this study. Thus, format differences do not account for the disparity between the Horvath (1) and Podlesny and Raskin (12) findings.

In a review of the empirical findings on the relative merits of Exclusive and Nonexclusive control questions, Krapohl, Stern, and Ryan (23) reported a position generally consistent with the findings here. However, they concluded, in part, with the following statement in favor of the continued application of Exclusive control questions: "…Therefore, nonexclusionary PLCs [Probable Lie Control Questions] may perform better in practice, despite the accepted wisdom of the exclusionary PLC… [However]…Polygraphers and agencies can rely instead on the nonscientific considerations cited earlier to continue a preference for this type [exclusionary] of comparison question" (p. 249).

Given these findings, the position offered by Krapohl, Stern, and Ryan (23) may have been diplomatic, but it is empirically unsupported. Only one report on this topic has revealed any advantage in the use of Exclusive control questions (12); that "advantage," by the way, had to do with better performance only on innocent persons, a finding that is contrary to all other reports. All the others including one field assessment, reported an advantage for Nonexclusive control questions (24).

The evidence regarding the two types of control questions examined here, based on the accumulated research findings, is not ambiguous. It is certainly sufficient to overcome the "nonscientific considerations" advanced by Krapohl, Stern, and Ryan (23) in support of a continuation of the "preferred" practice of using exclusionary control questions. It may be that Nonexclusive control questions are more difficult to develop than exclusive questions and they may require more training and skill to employ. Nevertheless, they reduce the likelihood of false-positive errors and they enhance the differentiation between guilty and innocent subjects; these are both realized without any loss in the detection of "deception" or an increase in false-negative errors. When these findings, resulting from a carefully structured laboratory study, are joined with other such studies (1,12) and with field data such as Amsel's (24) report, there is solid reason to discard the premise used to support Exclusive control questions.

That premise holds that when there is no overlap in coverage between the relevant questions and the control questions, an examinee differentiates better between the two question categories. For innocent persons, this means that the subject focuses more directly on the control questions and less on the relevant questions. For guilty persons, the opposite occurs and because there is no overlap, false-negative errors are minimized. For both subject groups, inconclusive outcomes are reduced. The results here, as well as in all other empirical reports, are strongly and directly contrary to this frequently advanced "common sense" argument (23). An alternate theoretical premise, one that is consistent with these and other findings is that Nonexclusive control questions are always less limited in scope of time and subject matter coverage; they are, therefore, more likely to be a "lie" or to cause doubt or concern about the accuracy of the subject's answer. It can also be said that from the subject's point of view, Nonexclusive control questions may be more difficult to distinguish from relevant questions. This, in turn, might suggest greater uncertainty about how responses to the control questions will be evaluated vis-à-vis the relevant questions. From this perspective, therefore, Nonexclusive control questions are more likely to produce better differentiation; these results support that hypothesis. Although the results on guilty persons were not statistically different for Exclusive and Nonexclusive control questions, there is no evidence that the relevant control question overlap produces either more inconclusive outcomes or more false-negative errors. The results on the innocent persons were statistically significant and rather pronounced in favor of Nonexclusive control questions. If the findings here generalize to the field, as suggested by Amsel's (24) results, innocent persons are better protected by the use of Nonexclusive control questions; false-positive errors are dramatically reduced with no disadvantage in detecting guilty subjects.

That CQ testing can discriminate between guilty and innocent persons involved in a specific event, as these findings show, is not a surprise. Such a finding is common and there are a large number of studies supporting that fact (4). A related point, though, is that these results were entirely consistent with those reported in other studies wherein, unlike here, the environment was ostensibly structured deliberately so as to enhance the potential for high accuracy (10). The use of college students, a relatively low motivation to succeed, and a relatively uninvolving scenario in this study did not produce a loss in accuracy from what has been reported when those factors have been deliberately avoided. However, it is to be noted here that relatively low accuracy in this study on innocent subjects when Exclusive control questions were used was considerably different from results in some other studies (1,7,12). An explanation for this discrepancy is not readily apparent. It is yet to be determined what design elements in experimental, laboratory-based research on CQ testing account for the often noted disparity in findings (4).

These results overall support the theory posited by advocates of CQ polygraph testing, that in the context of a specific event, innocent (truthful) and guilty (deceptive) subjects will respond differentially to relevant and properly developed control questions (9). The blind, numerical scoring of the polygraph charts here showed that this was the case. Moreover, the subjects' subjective expressions of their level of concern for relevant and control questions also varied according to Status; innocent subjects reported more concern for control questions than for relevant questions. Guilty subjects, in accordance with expectations, reported the opposite. These results taken together, the chart scorings and the subjective reports of the subjects, bolster the findings reported in earlier studies and the theory on which CQ testing is based as found in the field and scientific literature (1,9,15,25,26).

It might be said that CQ testing, as implied earlier in this paper, is to an unknown degree influenced by the preferences and practices of the examiner, the person carrying out the testing. Indeed, this is sometimes said to account for the relative lack of scientific interest in the CQ testing process (2,4). Thus one could argue that in this study the strong and consistent effects showing the benefit of Nonexclusive relative to Exclusive control questions were merely a reflection of the examiner's preferences. The examiner here, however, was trained in a facility that emphasized the use of the ZoC procedure with Exclusive control questions; the procedure and those question types were applied in his daily practices as a polygraph examiner for over 3 years. If there were an operational bias, it would appear to have been contrary to these findings.

Two of the variables examined in this study, subjects' gender and the testing format are of special interest. With respect to the latter variable, it has been the general perception among field examiners that the ZoC procedure, because it incorporates what are believed to be internal "protective" features such as a "symptomatic" question and a "sacrifice relevant" question, is to be preferred. The extant research that has isolated and addressed these specific points, however, has not been supportive. Neither the symptomatic question nor the sacrifice relevant question have been shown to function in a way that is consistent with that proposed by those who advocate their use; in other words, these features do not enhance the ZoC procedure relative to other approaches (11,13,14). In spite of that empirical evidence, the ZoC is still presumed to yield a higher accuracy than other procedures such as the MGQT. In fact, this issue has become a controversial one in the field

literature and as yet it is not resolved (27,28). However, the data on which the two approaches have been evaluated are cross-study findings; that is, there has been no scientific research in which the two procedures have been directly assessed under the same controlled conditions as carried out here. These findings show that when the two procedures are applied in identical conditions (and employing identical relevant questions) they do not yield statistically significant differences in accuracy or in the numerical scorings that are used for decision-making. Whether or not this finding generalizes to field settings, however, needs to be investigated.

One point worthy of attention with respect to the testing procedures is that when each was applied in this research three of the relevant questions were identical. When the other two questions used in the MGQT format were scored and those scores were included in the total, the accuracy (excluding inconclusives) of that procedure was slightly reduced relative to the three question version. This reduction, however, was modest and not statistically significant irrespective of the Type of control question.

Although there have been several studies of CQ testing in which males and females have been included, the effect of that variable has not typically been systematically evaluated. When it has been assessed it has not produced clear statistically significant results (20,29). Similarly, in this study, the gender of the subjects did not produce consistent effects on any of the dependent measures. However, there were two higher-order interaction effects involving subjects' gender. These were shown only when the numerical scorings of the individual physiological sensors (upper pneumograph and electrodermal) were analyzed. It is of interest to note that when the effect of subjects' gender was systematically manipulated in a study involving a variation of Information Recognition Testing (sometimes referred to as "Guilty Knowledge Testing," though there is no consensus on the essential elements of these procedures.) there was no statistical relationship to the accuracy of outcomes (18). Although that study did not involve "innocent" subjects, those findings considered together with these results suggest that subjects' gender is unrelated to the accuracy of commonly used polygraph testing procedures. Subjects' gender, however, may be associated with specific features in the physiological sensors, a finding suggested by these results and in the report of Timm (18).

In summary, these findings, in addition to confirming earlier results in both laboratory and field settings are contrary to the dogma that has been advanced in much of the polygraph literature in courtroom settings and in training facilities. These findings reveal clearly the pressing need for more and better empirical support for ideas that underlay accepted practices in polygraphy. The uncritical acceptance of seemingly authoritative and intuitively plausible "common sense" ideas, which are empirically unsupported, must be addressed directly if polygraphy is to progress in a meaningful scientific way.

References

1. Horvath F. The utility of control questions and the effects of two control question types in field polygraph techniques. J Police Sci Admin 1988;16(3):198–209.
2. Ben-Shakhar G. A critical review of the Control Question Test (CQT). In: Kleiner M, editor. Handbook of polygraph testing. San Diego, CA: Academic Press, 2002;103–26.
3. Iacono WG, Lykken DT. The scientific status of research on polygraph techniques: the case against polygraph tests. In: Faigman DL, Kaye DH, Saks MJ, Sanders J, editors. Modern scientific evidence: the law and science of expert testimony. St. Paul, MN: West Law, 2002;2: 483–538.
4. National Research Council. The polygraph and lie detection. Washington, DC: Committee to Review the Scientific Evidence on the Polygraph, Division of Behavioral and Social Sciences and Education, National Academies Press, 2003.
5. Honts CR, Raskin DC, Kircher JC. Scientific status: the case for polygraph tests. In: Faigman DL, Kaye DH, Saks MJ, Sanders J, editors. Modern scientific evidence: the law and science of expert testimony. Vol. 4, Forensics 2005–2006 ed. Thompson West: Eagan, MN, 2005; 571–605.
6. Reid J. A revised questioning technique in lie detection tests. J Crim Law Criminol 1947;37:542–7.
7. Barland GH, Raskin DC. An evaluation of field techniques in detection of deception. Psychophysiology 1975;3:321–30.
8. Barland GH, Raskin DC. Detection of deception. In: Prokasy WF, Raskin DC, editors. Electrodermal activity in psychological research. New York: Academic Press, 1975;417–77.
9. Reid J, Inbau F. Truth and deception: the polygraph (lie-detector) technique. Baltimore, MD: Williams & Wilkins, 1977.
10. Kircher JC, Horowitz SW, Raskin DC. Meta-analysis of mock crime studies of the control question polygraph technique. Law Hum Behav 1988;12:79–90.
11. Barland GH. A dual issue format. Polygraph 1983;1:1–6.
12. Podlesny JA, Raskin DC. Effectiveness of techniques and physiological measures in the detection of deception. Psychophysiology 1978;4:344–59.
13. Horvath F. The value and effectiveness of the sacrifice relevant question: an empirical assessment. Polygraph 1994;4:261–79.
14. Krapohl DJ, Ryan AH. A belated look at symptomatic questions. Polygraph 2001;3:206–11.
15. Honts C, Amato S, Gordon A. Effects of outside issues on the comparison question test. J Gen Psychol 2004;1:53–74.
16. *United States v. D. F. Jameson*. Correctional holding detachment, U.S. Disciplinary Barracks, Court Martial Convening Order Number 3. Fort Leavenworth, KS: Headquarters, Department of the Army, Combined Arms Center, 1988.
17. *U.S. v. R. C. Williams*. I:03-CR-636-5-JEC, Atlanta, GA: U. S. District Court for the Northern District of Georgia, 2005.
18. Timm H. Analyzing deception from respiration patterns. J Police Sci Admin 1982;10:47–51.
19. Pollina DA, Dollins AB, Senter SM, Krapohl DJ, Ryan AH. Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. J Appl Psychol 2004;89–6:6, 1099–105.
20. Honts CR, Hodes RL, Raskin DC. Effects of physical countermeasures on the physiological detection of deception. J Appl Psychol 1985;1:177–87.
21. Patrick CJ, Iacono WG. Validity and reliability of the control question polygraph test: a scientific investigation [Abstract]. Psychophysiology 1987;24:604–5.
22. Lykken DT. A tremor in the blood: uses and abuses of the lie detector. New York, NY: Plenum Trade, 1998.
23. Krapohl DJ, Stern BA, Ryan AH. Exclusionary or non-exclusionary: a review of the evidence. Polygraph 2003;4:245–50.
24. Amsel T. Exclusive or non-exclusive comparison questions: a comparative field study. Polygraph 1999;4:273–83.
25. Bradley MT, Janisse MP. Accuracy demonstrations, threat and the detection of deception: cardiovascular, electrodermal, and pupillary measures. Psychophysiology 1981;3:307–15.
26. Offe H, Offe S. The comparison question test: does it work and if so how? Law Hum Behav 2007;31:291–303.
27. Krapohl D. Validated polygraph techniques. Polygraph 2006;35-3:149–55.
28. Krapohl D. Rejoinder to criticisms by Matte: closing comments. Polygraph 2007;31-1:48–54.
29. Honts CR, Raskin DC, Kircher JC. Detection of deception: effectiveness of physical countermeasures under high motivation conditions [Abstract]. Psychophysiology 1983;20:446–7.

Additional information and reprint requests:
Frank Horvath, Ph.D.
Emeritus Professor
School of Criminal Justice
Michigan State University
108, Columbia Club
Blythewood, SC 29016
E-mail: Horvath@msu.edu