# (U) Credibility Assessment Standardized Evaluation (CASE) Challenge

Alexis Jeannotte, Ph.D.
Program Manager

**July 31, 2018**

INTELLIGENCE ADVANCED RESEARCH PROJECTS ACTIVITY (IARPA)

# Agenda

- Disclaimers/Ground Rules

- Goals

- Challenge Background

- Challenge Details

- Challenge Participation

- Questions

- Terminology

- Recommended Reading

# Disclaimers

- This Participants' Day Webex
    - Is provided for information and planning purposes
    - Is **not** the formal launch of the challenge
    - Does not constitute a formal solicitation for a challenge, proposals, or proposal abstracts

- Nothing said during this Participants' Day Webex changes the challenge requirements detailed in the future prize challenge
    - Details are subject to change at challenge release

# Important Points

- The audio and slides will be posted to the IARPA.gov website.

- This Participants' Day is being recorded, so to eliminate background noise all participants will be muted during the overview.

- Please identify the chat window in your WebEx interface. There may be updates shown there throughout the meeting.
  - Please use the chat window to let us know if you can't hear the presenter or see the slides

# Questions & Answers

- We'll have breaks for questions.

- Please send us your questions:
  - Email us your questions throughout the talk to:

    **CASEChallenge@iarpa.gov**

    with the subject "Webex Question"

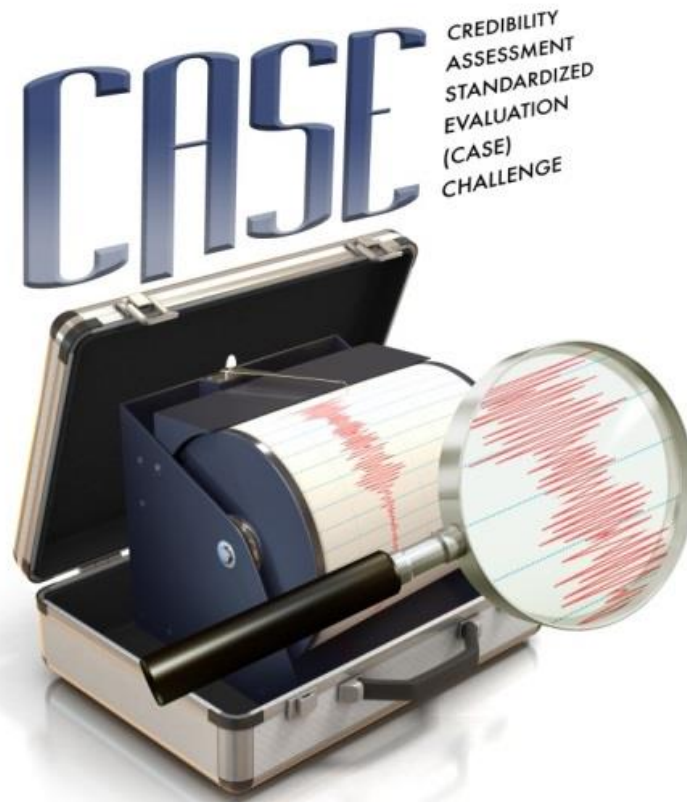  - Type the question in the chat window

# Participants' Day Goals

- Explain the vision, goals, and background for the expected challenge

- Provide information about the expected awards and scoring

- Invite questions and feedback to clarify the vision, goals, and background of the expected challenge

# CASE Challenge

**Email questions to: CASEChallenge@iarpa.gov**   UNCLASSIFIED   Intro | **Background** | Details | Participation

# CASE Challenge Overview

*The CASE Challenge will invite solvers to develop **novel, valid, reliable,** and **repeatable** ways to **measure** the performance of **credibility assessment** techniques and technologies.*

# CASE Challenge Overview

The CASE Challenge will invite solvers to develop **novel, valid, reliable,** and **repeatable** ways to **measure** the performance of **credibility assessment** techniques and technologies

- **Credibility assessment** refers to an assessment of many factors of a person and/or their information, to include truthfulness and trustworthiness

- Today, it is difficult to measure the accuracy of credibility assessment techniques, like the polygraph, using a standardized, repeatable method with good validity and reliability

- This challenge is focused on developing new methods to evaluate how well credibility assessment techniques work

- This challenge is **not** focused on developing new techniques, nor testing existing credibility assessment techniques

# CASE is NOT Focused on the Credibility Assessment Techniques or Technologies

- **Examples of credibility assessment techniques and technologies that are <u>NOT</u> the focus of the challenge**
  - Polygraph
  - Ocular movements and pupillometry
  - Detecting movement/fidget
  - Linguistic analysis
  - Questioning techniques to elicit information

Email questions to: CASEChallenge@iarpa.gov

Intro | **Background** | Details | Participation

# Who cares about credibility?

- Across our personal and professional lives we care about whether someone is reliable, honest, and trustworthy – the person, their actions, and/or what they say.

- Credibility assessments are used across a range of applications by federal, state, and local governments, as well as in the private sector
  - For example, the U.S. federal government carries out thousands of polygraph tests each year on job applicants and current employees (National Research Council, 2003)

- Credibility assessment applications, include:
  - Personnel screening
  - Criminal investigations
  - Source vetting
  - Others

# Everyday Uses of
# Credibility Assessment Techniques

- **Issue-Specific**
  - Questions tightly focused on specific acts (e.g. after a crime)
    - "Did you see the victim on Monday?"
  - Primarily retrospective questions
  - Often used in criminal and civil investigations
  - Most evaluation protocols (both laboratory & field) focus on issue-specific use cases

- **Screening**
  - Questions more generic, examinee may be uncertain about guilt or truth
    - "Do you intend to comply with your non-disclosure agreement?"
  - Both retrospective and prospective questions (i.e. future intentions)
  - Often used in pre-employment screening for trusted positions
  - Few evaluation protocols (both laboratory & field) focus on screening use cases

# Credibility Assessment: The State of the Science

There are a variety of ways organizations such as the US Government have tried to study and evaluate the credibility of individuals and/or their information

## Laboratory Protocols

- Conducted in well-controlled environments (not always explicitly in a laboratory!)
- Typically conducted for the purpose of
  - Generalizing the findings to real world uses or other populations
  - Understanding causality (not just correlation)

## Field Protocols

- Conducted in the natural location, time, and conditions under which credibility would normally be assessed
- Typically evaluate performance of a credibility assessment technique or technology under realistic conditions (e.g. nervousness, use of interpreter, environmental conditions)

Email questions to: CASEChallenge@iarpa.gov

Intro | **Background** | Details | Participation

# Evaluation Methods: Laboratory Protocols

## ■ **Examples**

- ☐ Committing or observing a mock crime (e.g. jewel theft, bomb scenario described below)

- ☐ Cheating or helping in cheating

- ☐ Breaking or causing dysfunction of an object

- ☐ Share or create an autobiographical experience or opinion (e.g. a time visiting a location, learning how to do a new activity)

"If you agree to participate in the study, you will be asked to take a lie detector test. You have been randomly selected to participate in a mock crime, and then lie about what you did. Prior to taking the lie detector test, you will be asked to go outside, locate a mock bomb, and place it next to a nearby road. During the lie detector test, your job will be to lie about the incident to the lie detector operator and convince the operator that you are telling the truth." (from Senter et al. 2006)

Senter, S., Waller, J., & Krapohl, D. (2006). Validation studies for the preliminary credibility assessment screening system (PCASS). *Department of Defense Polygraph Institute*.

# Evaluation Methods: Laboratory Protocols

**Sample Experiment**

> **Participants recruited, complete informed consent, and enrolled in study**

⬇

> **Participants randomly assigned to a group (e.g. be forthcoming or conceal information)**

⬇

> **Participants complete a set of actions that are defined as correct/not correct, good/bad, direct experience vs. observation, etc.**

⬇

> **Participants are asked about the set of actions to determine whether they are being credible or not credible**

⬇

> **Participants receive compensation, full details of the experiment are explained, and questions are answered**

Some experiments involve participants choosing whether they will be forthcoming or conceal information

Participants are typically assigned to one group, but in some experiments have chosen their group

Some experiments involve bonuses for participants if they are able to convince an interviewer that they are credible

# Evaluation Methods: Laboratory Protocol Limitations

- Low stakes and low perceived consequences if they are deemed not credible
  - And often associated with monetary incentives

- Generally lower stress and emotional reactivity

- Participant credibility is "instructed" (e.g. randomly assigned to lie or not lie)

- Laboratory protocols are artificially unambiguous (e.g. what constitutes lying is simple and transparent)

- Rate of participants that are credible/not credible do not reflect rates in real life

- Participant demographics often differ from populations in real life

Pollina et al. (2004) compared polygraph results between mock crimes and actual criminal investigations. Their findings suggest:
- Differences in psychological and physiological activity, as well as demographics of tested participants, might be responsible for differences in polygraph accuracy seen in mock crime protocols
- Metrics designed for evaluating an individual's credibility in real life may not work well in laboratory settings
- More motivated participants or more realistic mock crime scenarios may result in laboratory responses more comparable to the responses seen in real life

# Evaluation Methods: Field Protocols

- **Examples**

  - Participants are participating in a real law enforcement investigation

  - Participants are undergoing screenings for employment in trusted positions

  - Use of historical events (e.g. trials in court, public statements or filings, corporate communications [e.g. Enron])

  - Use of observational techniques (e.g. Yelp reviews, public social media posts)

Brownlie et al. (1998, unpublished manuscript) produced results looking at real screening situation: 769 tests of applicants for security positions at Atlanta International Airport between 1995 and 1997. The tests included four questions on past convictions for traffic violations or felonies, past bankruptcies, and use of an illegal substance during the past 30 days.

Brownlie, C., G.J. Johnson, and B. Knillant Screening Format. Unpublished manuscript. National Security Agency, Washington, DC.

**Email questions to: CASEChallenge@iarpa.gov**    Intro | **Background** | Details | Participation

# Evaluation Methods: Field Protocol Limitations

- **Natural tradeoff with lab studies**
  - Frequently lack a clear and independent determination of ground truth, so concerns about the accuracy of the information used to verify whether a person and/or their information is credible (e.g. participant confessions, background check, trial outcomes, etc. )
  - Impossible to control for all variables
  - Random assignment of participants to groups is not possible
  - Methods must be customized for the different "field conditions" so they will vary across experiments

- **Legally and logistically difficult to use with novel approaches**
  - Barriers to introducing novel approaches into current real life situations

# Summary of Existing Methods & Limitations

- To date, methods to evaluate credibility assessment techniques have:
  - Assessed credibility related to specific issues
  - Relatively fewer studies have looked at assessing credibility for screening applications
  - Been performed in lab settings and real-life (or field) environments

- Challenge participants may develop methods for specific issues or screening applications, the lab or field evaluation, but will need to develop methods that don't have the limitations of the current methods

# A Challenge to Address these Limitations

The CASE Challenge will invite solvers to develop **novel, valid, reliable,** and **repeatable** ways to **measure** the performance of **credibility assessment** techniques and technologies

Draw upon individuals with differing experiences and knowledge

Foster innovation through crowd-sourcing

Reduce barriers to entry with simplified, low tech solutions

Cultivate and sustain a collaborative community dedicated to this topic

**Email questions to: CASEChallenge@iarpa.gov**

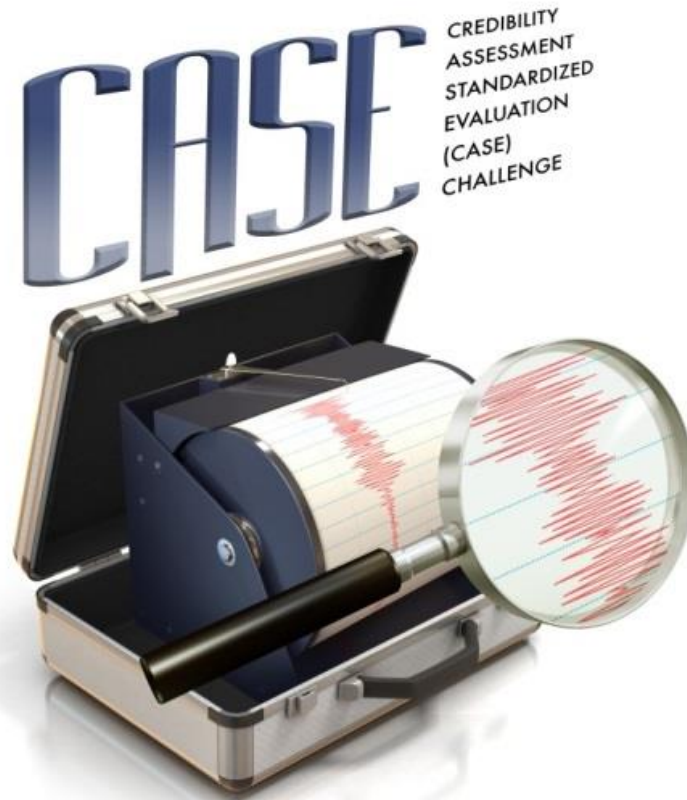Intro | **Background** | Details | Participation

# Successful Solutions will Motivate...

- **A new standard to use in evaluating current and future credibility assessment techniques and technologies**

- Use across Government, Academia, and Industry

- The development of next generation credibility assessment techniques

**Email questions to: CASEChallenge@iarpa.gov**

Office of the Director of National Intelligence
I A R P A
BE THE FUTURE

# CASE Challenge Details

**Email questions to: CASEChallenge@iarpa.gov**          UNCLASSIFIED          Intro | Background | **Details** | Participation

# Challenge Overview

- Solvers will develop protocols to measure the performance of credibility assessment techniques

- Solvers will submit their protocol in the form of a short paper

- Submissions will be evaluated by a panel of scientific experts according to a set of criteria

- Winners will be awarded prizes

We will **NOT** run experiments or collect data as part of the evaluation process

Office of the Director of National Intelligence
**I A R P A**
BE THE FUTURE

# CASE is Focused on the Protocol

**Protocol: Plan for conducting an experiment**

How is credibility defined?

Where does this take place?

What will participants be asked to do, think about, respond to, etc.? What instructions will they receive?

How will participants be motivated?

What are the positive/negative consequences if they are evaluated as credible or not credible?

What is/are the key piece(s) of information, action(s), decision(s), etc. that will be used to demonstrate a participant or their information is credible?

How does the experimenter objectively record, observe, or otherwise know the truth about a participant's credibility?

How do these steps reflect something real? Is it repeatable?

**Not the focus of the CASE challenge**

Credible
or
Not Credible?

Objective Truth About
Credible or Not Credible

**Email questions to: CASEChallenge@iarpa.gov**

Intro | Background | **Details** | Participation

# Solvers will develop and provide solutions that address...

- **Overview**
  - Provides summary, objectives and background information
- **Procedures**
  - Describes overall design and duration, including sequence and timing of actions or events and roles / responsibilities of all involved
  - Procedures must be repeatable by any capable researcher with protocol in hand and access to necessary resources
  - Objective truth needs to be attainable
  - Participants need to be motivated and have choices about when and/or how to be credible

- **Participant Risks**
  - Details any potential risks that could occur as a result of study participation and steps taken to minimize those risks
- **Real-World Applicability**
  - Describes how protocol represents real-life conditions / settings
- **Supplementary Information**
  - Provides relevant information not covered in other sections

*Exact submission format will be published at challenge release

# Sample "Mock Bomber" Protocol

| Participants | United States Army basic trainees at training site | |
|---|---|---|
| **Procedures** | "Not credible" group instructed to place a bomb, arm it, and if questioned, admit to knowledge of the bomb, but deny personal involvement. | **Issue-specific use case** |
| | Participants hear actual detonation after planting mock bomb. | **Close simulation of real-world conditions** |
| | "Credible" group told about the bomb, but instructed to go to a break area and cooperate if questioned about their knowledge of a bomb or their personal involvement. | **Control group** |
| | All participants told that if they are found to be lying they will have to give a public speech on honesty, integrity, and loyalty. Speech is unenforced. | **Establishes consequences** **Retrospective design** |
| **Credibility Assessment** | Is the participant assessed as credible or not credible?  **NOT PART OF SUBMISSION** | **Protocol is agnostic to the specific credibility assessment method** |
| **Results** | How often does the credibility assessment result match the ground truth? | **Straightforward to evaluate** |

Based on Senter, Stuart, James Waller, and Donald Krapohl. "Air Force Modified General Question Test Validation Study." Polygraph 37, no. 3 (2008): 174-184.

# What the CASE Challenge is NOT

- The CASE Challenge is **NOT** seeking new technologies or techniques for performing credibility assessment
  - The CASE Challenge is technology agnostic

CASE Challenge submissions **will be** in the form of written **PROTOCOLS** that can be used with any credibility assessment technique or technology

Submission templates will be provided when the challenge launches

# Submission Disqualifiers

- Submissions should not involve developing or modifying credibility assessment technologies/techniques, but should instead focus on creating novel methods for evaluating how well current and future credibility assessment techniques and technologies work

- Human subjects research studies cannot be conducted as part of the challenge

- Challenge participants cannot submit data from a human subjects research studies
  - Data will **NOT** be factored into the evaluations
  - However, Solvers may cite previous publications in their submission

- Solutions cannot include plagiarized materials

- Challenge participants cannot submit materials for which they do not own the rights

# Challenge Scoring Overview

- Evaluators will not know who submitted a solution

- Each protocol will be evaluated separately

  - Solvers may submit multiple solutions

  - Each solution will be evaluated independently of one another

- Protocols will be evaluated based on the text of the submission and not on the results from previously conducted research. *(Please Remember: Human subjects research will not be conducted as part of this challenge)*

Email questions to: CASEChallenge@iarpa.gov

# Challenge Scoring Details

- CASE Challenge protocol submission scoring will consist of two phases

- Phase 1: Compliance Review:
  - Conducts an initial screening of submissions
  - Submissions that do not comply with specified challenge requirements will not be further evaluated

- Phase 2: Scoring Panel:
  - Members of the scoring panel individually score each submission that passes compliance review
  - Scores are based on a set of predefined evaluation criteria

Email questions to: CASEChallenge@iarpa.gov                    Intro | Background | Details | Participation

# Sample Challenge Scoring Criteria

| Criteria | Short Description |
|----------|-------------------|
| Scientific Credibility | Able to generate valid, reliable and repeatable outcomes |
| Real-World Applicability | Accounts for real life conditions and settings |
| Novelty | New and original, as compared to existing protocols |
| Participant Considerations | Safe and ethical, adheres to human subjects protection guidelines |

*Exact point values and specific criteria will be conveyed at challenge release

# Prize Purse

- Up to $100,000 will be awarded in prize money

- Selected winners will also be given a travel award to attend a challenge workshop that will bring together people from academia, the government, the private sector, and others who are interested in this topic.

Email questions to: CASEChallenge@iarpa.gov

Intro | Background | Details | Participation

# Note about Intellectual Property

- It is the Solver's responsibility to ensure they have the right to distribute the IP.

- Challenge participants will be asked to sign a form stating they have the rights to the IP as part of their registration or submission process for the challenge.

- It is expected that solvers will make solutions publicly available.

# CASE Challenge Participation
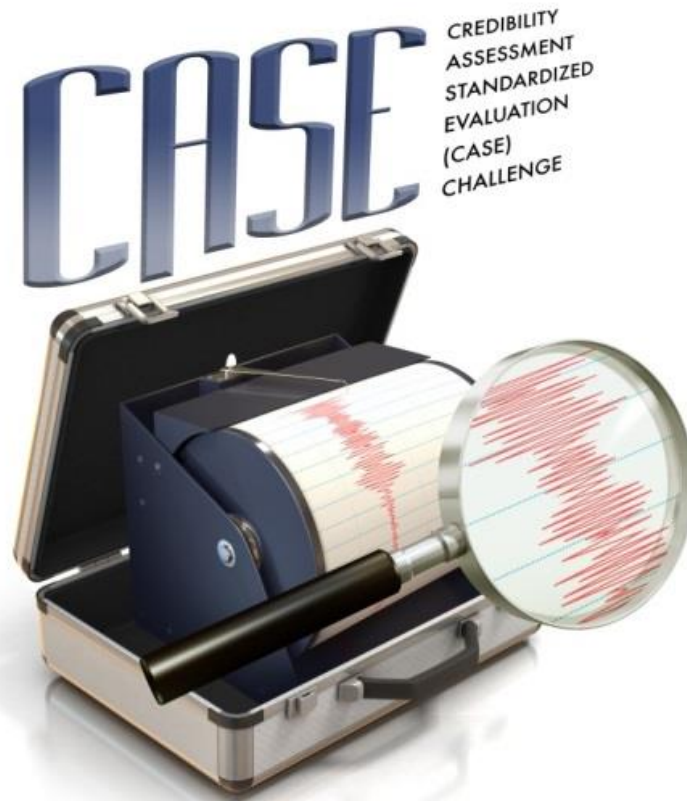
# Eligibility to Participate

- In order to be eligible to participate you must:
  - Be over 18 years of age
  - Submit protocol(s) in clear, grammatically correct English

- Individuals from around the world welcome to participate
  - Diversity in thought, experience and background is encouraged

- The CASE Challenge is looking for solutions from anyone who thinks they might have a way of addressing the credibility assessment evaluation problem
  - This includes individuals in fields outside of those normally involved in this type of research

Email questions to: CASEChallenge@iarpa.gov

Intro | Background | Details | Participation

# CASE Challenge Timeline

| Challenge Event | Tentative Dates |
|---|---|
| Virtual Participants' Day | TODAY |
| Challenge Release | Fall 2018 |
| Submission Deadline | Winter/Spring 2019 |
| Evaluation Close | Spring 2019 |
| Workshop | Summer 2019 |

**Email questions to: CASEChallenge@iarpa.gov**

# Partner Organizations

- ## National Center for Credibility Assessment (NCCA)
  - Federal center for credibility assessment education, oversight, research and development
  - Challenge input and guidance

- ## Johns Hopkins University Applied Physics Laboratory (JHU/APL)
  - University Affiliated Research Center (UARC)
  - Challenge input and management

Email questions to: CASEChallenge@iarpa.gov                    Intro | Background | Details | Participation

# Frequently Asked Questions

- Where can I go to learn more about the CASE Challenge?
  - https://www.iarpa.gov/index.php/working-with-iarpa/prize-challenges/1119-credibility-assessment-standardized-evaluation-case-prize-challenge

- How will I submit my completed protocol(s)?
  - Protocol submission instructions will be provided at Challenge launch.

- Is there a limit on the number of protocols I can submit?
  - It is not expected that there will be a limit on the number of unique protocols one solver can submit.

- Will there be access to any IARPA data or experts while developing our protocol(s)?
  - No.

Email questions to: CASEChallenge@iarpa.gov

Intro | Background | Details | Participation

# Frequently Asked Questions (continued)

- Is there a preference for field-based protocols over lab-based protocols?
    - No, both field and lab protocols are of interest.

- Can the protocol include actions that happen in virtual spaces?
    - Yes.

- Should protocol submissions be limited to interview-based techniques for credibility assessment, such as is the case with polygraph examinations?
    - No. The CASE Challenge is interested in protocols that could be used to evaluate any credibility assessment techniques / technologies – even ones that haven't been developed yet!

## More FAQs to come – help us develop them!
## Please submit questions to our email address.

# Questions
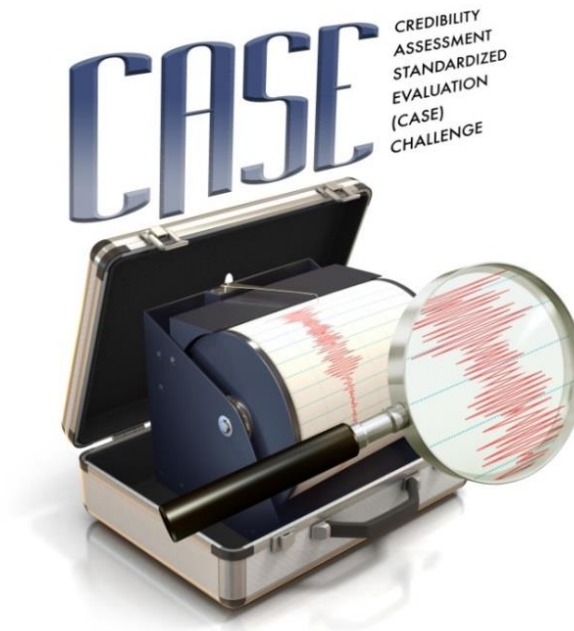


For questions that arise after today, please email:
Program Manager: **Alexis.Jeannotte@iarpa.gov**
Challenge Team: **CASEChallenge@iarpa.gov**

# Terminology

| | |
|---|---|
| **Novel** | New and original, as compared to existing protocols |
| **Valid** | Generates outcomes that measure what they claim to measure and that accurately correspond to real-world phenomena |
| **Reliable** | Generates consistent outcomes across a range of conditions, across subjects |
| **Repeatable** | Generates consistent outcomes under identical conditions |
| **Credible** | Person and/or information that is truthful (or true), trustworthy, and/or believable |

# Terminology Cont.

| | |
|---|---|
| **Measure** | A standard of comparison (in this case, as related to the performance of various credibility assessment techniques) |
| **Polygraph** | A device which records physiological measurements while a subject responds to questions |
| **Protocol** | A detailed plan for conducting an experiment (in this case, one which involves human participants) |
| **Ground Truth** | The objective real-world truth (in this case, as related to a specific claim or claims, or an individual's credibility) |

# Recommended Reading

- National Research Council. (2003). *The polygraph and lie detection*. National Academies Press. https://www.nap.edu/catalog/10420/the-polygraph-and-lie-detection

- Intelligence Science Board (2006). Educing Information: Interrogation: Science and Art, Intelligence Science Board, Phase 1 Report. https://fas.org/irp/dni/educing.pdf

- Intelligence Science Board (2009). Intelligence Interviewing: Teaching Papers and Case Studies, A Report from the Study on Educing Information. https://fas.org/irp/dni/isb/interview.pdf

- Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., & Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *Journal of Applied Psychology*, *89*(6), 1099. http://psycnet.apa.org/record/2004-21169-016

- American Psychological Association. (2012). Guidelines for ethical conduct of behavioral projects involving human participants by high school students. http://www.apa.org/science/leadership/research/ethical-conduct-humans.aspx

*An annotated bibliography will be provided at challenge release