Pacific Northwest
NATIONAL LABORATORY

PNNL-38983

Proposed Critical Thinking Metrics for Power Grid Trainees

January 2026

Michelle Dowling
Corey Fallon
Eric Andersen

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-38983

Proposed Critical Thinking Metrics for Power Grid Trainees

January 2026

Michelle Dowling
Corey Fallon
Eric Andersen

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

The TRS Intro to Critical Thinking for Operational Decision-Making course for power grid operators seeks to help trainees improve their use of critical thinking skills in operation environments. To determine the effectiveness of this training, we have identified 3 categories of evaluation measures: critical thinking metrics, situation awareness (SA) evaluations, and behavioral observations. Critical thinking metrics seek to directly measure trainee's ability to use critical thinking skills through their verbalized responses to different questions or scenarios. SA will be evaluated during various points throughout simulated exercises completed during the training, which will be hosted in Pacific Northwest National Lab's (PNNL's) Electricity Infrastructure Operations Center (EIOC) in Richland, WA. SA evaluations aim to determine the level of a trainee's ability to perceive, comprehend and anticipate events in the operational environment. This construct will be assessed through verbalized responses to questions informed by the Situation Awareness Global Assessment Technique (SAGAT). Behavioral observations capture how trainees use the system, including which simulator displays or windows they use or actions they perform and in what order. These observations help us understand how the critical thinking training may inform operator action. We hypothesize that while these measures are correlated with each other, they provide different perspectives on each trainee's ability to gain SA and use critical thinking skills in operation environments. By comparing these measures in a pre-assessment and a post-assessment, we can determine how much improvement a trainee gains through the training. These measures may also be used during the training as appropriate. This document specifies these 3 categories of measures, which components of these measures currently seem to be the best candidates to use in the training course, and which parts of the training course is related to each component.

Acknowledgments

We would like to thank Mike Cassiadoro from Total Reliability Solutions (TRS), an instructor for power grid operation training courses who acted as our subject matter expert, for providing guidance and insight on appropriate critical thinking metrics to use in a new critical thinking training course.

Acronyms and Abbreviations

BA – Balance Authority
EIOC – Electricity Infrastructure Operations Center
PNNL – Pacific Northwest National Lab
RAS – Remedial Action Schemes
RC – Reliability Coordinator
SA – Situation Awareness
SAGAT – Situation Awareness Global Assessment Technique
TOP – Transmission Operator
TRS – Total Reliability Solutions

Contents

Figures

Tables

1.0 Introduction

The TRS Intro to Critical Thinking for Operational Decision-Making course for power grid operators seeks to help trainees improve their use of critical thinking skills in operation environments. In support of this goal, we will employ three different categories of evaluation measures to evaluate training effectiveness and gain insights into operator cognition and behavior: critical thinking metrics, situation awareness (SA) evaluations, and behavioral observations.

The critical thinking metrics focus on how the operators process information to build situation awareness, analyze the situation and information provided to them, and decide on the correct course of action. These metrics are a direct assessment of whether trainees learned the various techniques taught in the course. In addition to the critical thinking metrics, operator SA is also important to evaluate. Part of the value of operator critical thinking is its role in helping operators build SA. Therefore, we also include SA evaluations to determine the level of a trainee's SA. This construct will be assessed through verbalized responses to questions informed by the Situation Awareness Global Assessment Technique (SAGAT). Finally, behavioral observations will be assessed based on the steps operators executed in response to events during a simulated exercise. These metrics capture how critical thinking is reflected by operator action (e.g., which simulator displays or windows of the system were used or actions performed in which order). We expect all three categories of metrics to be positively correlated. For example, operators who engage in critical thinking will develop good SA and take the correct course of action. However, each of these measures captures a different facet of SA and critical thinking ability. Incorporating all three categories provides a comprehensive assessment of the training's impact on operator cognition and performance. Metrics may be compared between a pre-assessment and a post-assessment as well as across operator roles and learning activities within the training.

Training will include simulated exercises hosted in PNNL's EIOC in Richland, WA ("Shaping the Future Electric Grid") that allow operators to practice various critical thinking tools and techniques. During each exercise, the operators will respond to events introduced in the scenario, and their actions in response to these events will be recorded to capture behavioral observations. After each event, the research team will record the course of action taken by each operator to determine the efficiency and accuracy of their response to the event. Next, the instructor will conduct a debrief with the operators to discuss the event and their process for managing it. It is during this debrief that the instructor will ask the SAGAT style questions to assess SA as well as questions designed to measure critical thinking (see Figure 1).

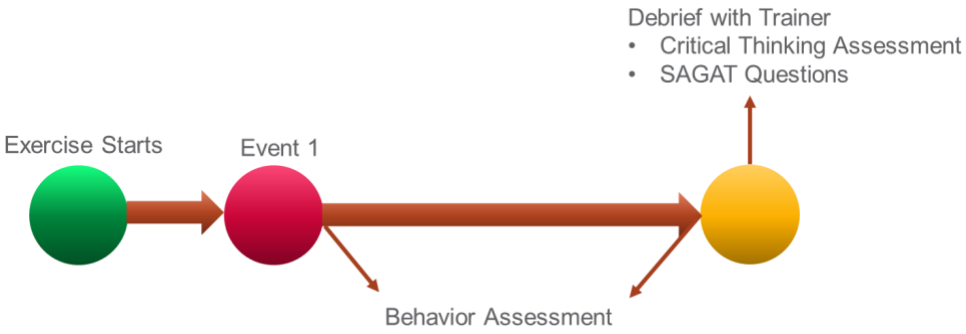

Figure 1. An illustration of when each type of human factors assessment will be collected during the simulated exercise. During debriefs with the trainer, SAGAT questions and critical thinking assessments will vary depending on the specifics of the scenario

2.0 Critical Thinking Metrics

To identify metrics that could measure how well a power grid trainee applies critical thinking skills in a learning activity or simulated environment, we first compiled a list of 14 different evaluations of critical thinking generated from those referenced in (Dunn et al. 2009, 72 – 75; Fero et al. 2010). Of those, 6 were outdated and therefore no longer had materials readily available for reference. For the remaining 8 evaluations (Fero et al. 2010; "California Critical Thinking" 2025; "Critical Thinking Skills Test" 2025; "Watson-Glaser Critical Thinking Appraisal III" 2020; Ennis and Millman 2005; "Cambridge International AS & A Level Thinking Skills 9694 Syllabus" 2020; Werner 1991; "The International Critical Thinking Essay Test" 2019; "Test of Everyday Reasoning" 2026), we extracted the different indicators of critical thinking that were included in the evaluation and counted how many of the 9 evaluations used each indicator. We also clustered the indicators using an affinity diagramming approach.

There are 3 main steps identified through this process for critical thinking: deconstructing information, analyzing a conclusion or proposed action, and forming a conclusion. In our discussions with a subject matter expert who is an experienced instructor for power grid operator training courses, we added an additional step at the beginning of the process: recognizing the need to think critically. For each step, we enumerate the recommended indicators for evaluating power grid operators who take the TRS Intro to Critical Thinking for Operational Decision-Making course. These indicators were selected as the most appropriate for capturing how well trainees apply critical thinking skills during the learning activities and simulated environments in the training. Each indicator has at least one associated metric. Metrics were similarly determined through a combination of research, input from a subject matter expert, and our own human factors expertise.

We propose the following indicators as the set to incorporate into this training course.

- Identifying the Need to Apply Critical Thinking Tools/Techniques
- Identifying Information
- Identifying Other Possibilities
- Inductive Reasoning
- Deductive Reasoning
- Abductive Reasoning
- Identifying Assumptions

The associated metrics for these indicators are designed to assess how effectively operators apply the critical thinking techniques taught in the training. These metrics allow us to evaluate operator critical thinking throughout the training such as between pre and post assessments. Our list of critical thinking metrics is considered a list of possible metrics. Final decisions on which subset of the proposed metrics to incorporate in the training course will be based on the constraints of the training and consultation with our TRS instructor. In addition, not all metrics will be applied across all aspects of the training. Each metric's relevance varies depending on the segment of the training. For reference, the current version of the training has been organized into 8 segments (see below).

- Segment 1: Modes of Thinking
- Segment 2: Functional Roles and Responsibilities
- Segment 3: Critical Thinking Framework

- Segment 4: Recognize the Need to Think Critically
- Segment 5: Assess the Situation
- Segment 6: Draw Conclusions
- Segment 7: Decide a Course of Action
- Segment 8: Final Case Study

It should be noted that there are other indicators of critical thinking that have not been included in our set of indicators above. While we believe these indicators such as identifying logical fallacies and cognitive biases are important for assessing critical thinking, these indicators are less relevant for our upcoming training. Descriptions of these indicators are included in Appendix A.

2.1 Metric for Recognizing the Need to Think Critically

The first step in critical thinking is recognizing the need to think critically. Without this step, it is likely that improper action will be taken based on automatic thinking (i.e., System 1 thinking).

2.1.1 Identifying the Need to Apply Critical Thinking Tools/Techniques

Is the power grid operator able to identify when they need to think critically?

Answering this question focuses on simply whether the power grid operator can **recognize the need to engage in manual thinking (i.e., Critical Thinking)** when prompted with a course of action, conclusion, or scenario that can benefit from critical thinking. Trainees reflect their ability to recognize the need to apply critical thinking through one of the following triggers:

1. Detection of abnormal conditions or unacceptable system performance that, if not addressed, could have an adverse impact on safety, reliability, or economics
2. Implementation of complex processes or procedures to achieve an operational goal
3. Realization that operational decisions are being made in an irrational or biased manner.

This metric aligns with Segment 4 of the training course as the segment focuses on how to identify the need to think critically.

Suggested Metrics:

- Percent of correctly identified prompts where critical thinking is needed:
$$\frac{number\ of\ correctly\ identified\ prompts}{total\ number\ of\ prompts} \times 100$$
    - For this metric, a *prompt* is any a course of action, conclusion, or scenario that can benefit from critical thinking
    - A correctly identified prompt is demonstrated by the trainer taking one of the three triggers listed above
    - This metric can be computed for a specific Segment or simulated exercise in the training or can be computed for an entire training course

2.2 Metric for Deconstructing Information

The next step in critical thinking is deconstructing information, which focuses on simply identifying the unfolding scenario and the key information needed to understand the scenario, without imposing any interpretation or additional information onto the scenario or key information. These metrics align with the "Assess the Situation" phase of the critical thinking framework provided in the training course, which is covered in Segment 5.

2.2.1 Identifying Information

How effective is the power grid operator at identifying all of the key information related to the problem at hand?

Answering this question leads us to measure the amount of **relevant information the power grid operator identifies** as they undergo their critical thinking process. In the training course, this relevant information takes the form of the key characteristics of the operational event presented. Especially for novice power grid operators, **how well they understand those characteristics or pieces of information** may also come into play. For poorly understood information, power grid operators may need to look for additional information or else may be subject to using (or ignoring) the information inappropriately. Of course, there is the possibility that the power grid operator **identified extraneous information** or **failed to identify relevant information**, so these are important elements to consider measuring as well.

This metric aligns with Segment 5 of the training course as the segment focuses on collecting information to gain clarity of the situation. Additionally, this metric aligns with Perception in SA.

Suggested Metric:

- Percent of relevant information that was missed: $\frac{amount\ of\ relevant\ pieces\ of\ infomration\ missed}{total\ amount\ of\ relevant\ pieces\ of\ information} \times 100$
    - Example types of information include:
        - Referenced
        - System Load(s)
        - Generation
        - Protection Schemes/RAS
        - System Outages
        - Interchange
        - Ratings/Limitations
        - Phase Angles
        - System Frequency
        - Interconnected Status
        - Abnormalities

Additional Potential Metrics:

- Percent of identified information that was extraneous: $\frac{amount\ of\ extraneous\ pieces\ of\ information}{total\ amount\ of\ information\ identified} \times 100$

- Understanding of identified relevant information
    - Individual piece of information
    - All pieces of information (likely an aggregate of the individual pieces)
- Relevance of each piece of information identified

2.3 Metrics for Analyzing a Conclusion or Action

Once the key information related to an event is identified a conclusion can be formed. The main component of critical thinking is analyzing this conclusion. In the context of power grid operations, the conclusion may take the form of a course of action. Analyzing a conclusion or action takes the key information identified in the previous step as well as the conclusion or action itself to determine which pieces of information are being used appropriately to support the conclusion or action.

If the conclusion or action is generated and shared by another operator, another key component to this analysis is considering the context in which the conclusion or action is provided. This context includes the point of view and reasons of the person providing the conclusion. Regardless of whether the conclusion was self-generated or constructed by another operator, an assessment must be made on the plausibility of the conclusion or action. This assessment is performed by evaluating the credibility of the key information in support of the conclusion or action.

Taken together, these components of analysis reflect the trainer's ability to consider the magnitude and scope of the event (which is aligned with Comprehension in SA) before they can determine the actions that must be taken to resolve the issue at hand and limit the potential impact on safety, reliability, and economics (with is aligned with Projection in SA).

2.3.1 Identifying Point of View, Reasons, and Possibilities

When evaluating a conclusion generated by someone else, the point of view and reasons for their conclusion provide critical context to understand why the information they provide is important or why the action is being requested. However, this additional context does not mean that the information or requested action is correct. The operator should think critically about the conclusion or action by considering other possible conclusions.

2.3.1.1 Identifying Point of View

What is the point of view of the person providing information or requesting an action?

Answering this question involves **understanding the perspective** of the person providing information or requesting an action.

This metric aligns with Segment 2 and Segment 5 of the training course as the segment focuses on the various functions and responsibilities of power grid operators. This metric aligns with the "Assess the Situation" phase of the critical thinking framework provided in the training course.

Suggested Metric:

- Percent of functions (i.e., Reliability Coordinator (RC), Balance Authority (BA), or Transmission Operator (TOP)) identified that matched the person providing information or perspective: $\frac{number\ of\ functions\ correctly\ identified}{total\ number\ of\ people\ providing\ info} \times 100$

2.3.1.2 Understanding the Reasons

What is the reason the person has for providing information or requesting an action?

Answering this question involves **understanding the motivation and reason** for the person providing information or requesting an action.

This metric aligns with Segment 2 and Segment 5 of the training course as the segment focuses on the various functions and responsibilities of power grid operators. This metric aligns with the "Assess the Situation" phase of the critical thinking framework provided in the training course, particularly the "Identifying Causes and Contributing Factors" technique.

Suggested Metric:

- Percent of motives identified that matched the relevant function (i.e., RC, BA, or TOP): $\frac{number\ of\ reasons\ that\ mached\ the\ relevant\ function}{total\ number\ of\ reasons} \times 100$

2.3.1.3 Identifying Other Possibilities

What other possible conclusions or actions could be made with similar information?

Identifying such alternative conclusions or actions opens avenues to explore, which may lead to identifying a better option. At this stage, simply **listing other possible conclusions or actions** is sufficient; these (as well as the originally proposed conclusion or action) will be evaluated in more detail through the metrics described in the following sections.

This metric aligns with Segment 6 of the training course as the segment includes instruction on considering alternative options while developing an action plan. Additionally, this metric aligns with Comprehension and Projection in SA.

Suggested Metric:

- Number of other possible conclusions or actions identified (count)

2.3.2 Methods of Reasoning

When assessing possible scenarios or courses of action, evaluating how key information are composed in support of the scenario or course of action can reveal its strength. For example, evaluating key information can illuminate faulty logic, cognitive biases, or improper assumptions (as discussed in the next section). Being proficient in a variety of methods of reasoning and identifying assumptions strengthen critical thinking capabilities.

2.3.2.1 Inductive Reasoning

Given specific statements, what generalizations can be made?

Inductive reasoning is a key method of reasoning that permits **making connections from specific instances to generalizations**. For example, inductive reasoning can connect different topics, points, or key information that may seem unrelated at a glance. Such connections can be critical to drawing a clear conclusion.

This metric aligns with Segment 6 of the training course as the segment includes instruction on drawing conclusions, including different methods of reasoning such as inductive reasoning and the "Differentiate Between Facts, Direct Experiences and Secondhand Information" critical thinking technique.

Suggested Metric:

- Percent of correct conclusions correctly drawn with inductive reasoning: $\frac{number\ of\ correct\ conclusions}{total\ number\ of\ prompts} \times 100$
    - For this metric, a *prompt* is any instance in the training requiring inductive reasoning
    - Prompts should take the form of: specific example 1, specific example 2; trainee provides generalized conclusion that can be drawn from those examples (Newstead et al. 2004; Leighton 2006)
    - This metric can be computed for prompts across a specific Segment or simulated exercise in the training or can be computed for an entire training course

2.3.2.2 Deductive Reasoning

Given general statements, what specific conclusions can be drawn?

Deductive reasoning is another key method of reasoning that permits **connecting concepts or higher-level ideas to draw specific conclusions**. Such connections can be critical to drawing a clear conclusion.

This metric aligns with Segment 6 of the training course as the segment includes instruction on drawing conclusions, including different methods of reasoning such as deductive reasoning and the "Differentiate Between Facts, Direct Experiences and Secondhand Information" critical thinking technique.

Suggested Metric:

- Percent of correct conclusions drawn with deductive reasoning: $\frac{number\ of\ correct\ conclusions}{total\ number\ of\ prompts} \times 100$
    - For this metric, a *prompt* is any instance in the training requiring deductive reasoning
    - Prompts should take the form of syllogisms (statement, related statement; trainee provides conclusion that can be drawn as a result of combining the statements) (Newstead et al. 2004; Leighton 2006)

- This metric can be computed for prompts across a specific Segment or simulated exercise in the training or can be computed for an entire training course

2.3.2.3 Abductive Reasoning

Given pieces of evidence, what conclusions can be drawn?

Abductive reasoning is another key method of reasoning that permits **following pieces of evidence to find the most probable explanation**. Such connections can be critical to drawing a clear conclusion.

This metric aligns with Segment 6 of the training course as the segment includes instruction on drawing conclusions, including different methods of reasoning such as abductive reasoning and the "Differentiate Between Facts, Direct Experiences and Secondhand Information" critical thinking technique.

Suggested Metric:

- Percent of correct conclusions drawn with abductive reasoning: $\frac{number\ of\ correct\ conclusions}{total\ number\ of\ prompts} \times 100$
  - For this metric, a *prompt* is any instance in the training requiring abductive reasoning
  - Prompts should take the form of: statement, unconnected or indirectly connected statement; trainee provides probable conclusion that can be drawn as a result of using those statements (Newstead et al. 2004; Leighton 2006)
  - This metric can be computed for prompts across a specific Segment or simulated exercise in the training or can be computed for an entire training course

2.3.2.4 Identifying Assumptions

What are the underlying assumptions of a conclusion or course of action, or the key information in support of that conclusion or action?

Assumptions reflect assumed context but can be inaccurate. **Identifying assumptions** and questioning the underlying beliefs highlight possible inaccuracies in logic, which should be clarified or corrected to draw a clear conclusion.

This metric aligns with Segments 6 of the training course as the segment includes instruction on drawing conclusions, including how to identify and test assumptions.

Suggested Metric:

- Percent of assumptions identified from prompts: $\frac{number\ of\ assumptions\ identified}{total\ number\ of\ assumptions} \times 100$
  - This metric can be computed for individual conclusions or courses of action that contain assumptions, across multiple conclusions in a specific Segment or simulated exercise in the training or for an entire training course

3.0 Situation Awareness Metrics

Operators' SA will be measured throughout the training scenario using SAGAT style questions (Endsley 2021). Our SAGAT style questions were designed to measure Endsley's three levels of situation awareness (Endsley 2017). Specifically, these questions evaluated operator perception (Level 1), comprehension (Level 2) and ability to project (Level 3).

- Perception – Level 1 SA: Perception of the status of elements in the operational environment.
- Comprehension – Level 2 SA: Comprehension is the integration of these elements to form an understanding of the operational environment in the context of task objectives.
- Projection – Level 3 SA: Projection is a prediction of the future state of elements in the operational environment based on one's comprehension of the current operational environment.

See Table 1 for example questions from each level of SA.

Table 1. Example SAGAT Questions for Each SA Level

SA Level	Example SAGAT Style Questions
Perception (Level 1)	What environmental elements changed and what key events occurred during the simulation?
Comprehension (Level 2)	How would you describe the current state of the power system? Please identify any critical issues that should be addressed in order.
Projection (Level 3)	Given the status of the grid and prevailing system conditions, how do you expect the system state to change as the day progresses?

Similar to the critical thinking assessment, SAGAT questions will be asked following specific time points throughout the simulated exercises. All participating operators will be asked the same set of SAGAT questions and responses will be rated by the instructor. Similar to previous onsite training conducting in the Electricity Infrastructure Operations Center (EIOC) that measured operator SA, SA ratings will range from 1 to 5 based on the amount of situation awareness exhibited in each response.

4.0 Operator Behavior Metrics

Although it is important to directly assess critical thinking and SA, these measures assess cognition, not behavior. We suspect operator critical thinking and SA will correlate with operator action. However, there are scenarios where behavior and cognition may not align. For example, it is possible operators may know the correct course of action in response to an event but have difficulty articulating their understanding of the environment and course of action. We believe that in addition to our cognitive assessments, it is also important to measure how the training translates into operator action.

During our simulation-based exercises the training team will insert events that operators must respond to. During these periods that the team will record operators' course of action (see Figure 1). Operator course of action can be constructed from the following sources of data.

- Operator key strokes and mouse clicks automatically logged during the simulated exercises
- Simulated grid responses to user interactions automatically recorded by the EIOC's PRIME testbed
- Screen recordings of operator interactions

If necessary, these data will by synched via time stamps to create a complete picture of operator behavior in response to simulated events.

From an operational perspective, the most valuable data to collect is the order of displays viewed as it provides insight into how the operator builds situation awareness and allows the observer to recognize the operating characteristics (e.g., generation output, voltage, power flow, equipment status, etc.) that the operator considers most important given the situation at hand. Recording this order allows us to identify gaps or extra steps in their process that might suggest incomplete situation awareness and/or challenges applying the critical thinking techniques covered in the training.

For each simulated event, operator course of action will be compared to an optimal baseline course of action. Quantitative metrics will be calculated to capture the accuracy of the operators' course of action.

Suggested Metrics: Content Accuracy

True Positive (TP) = Number of steps that appear in both the Observed Course of Action (O) and the baseline (B).

False Positive (FP) = Number of extra steps in O that are not in B

False Negative (FN) = Number of missing steps in O that are in B

Content Accuracy $= \frac{TP}{TP+FP+FN}$

- Content accuracy: $\frac{TP}{TP+FP+FN}$

  - True Positive (TP) = Number of steps that appear in both the Observed Course of Action (O) and the baseline (B).
  - False Positive (FP) = Number of extra steps in O that are not in B
  - False Negative (FN) = Number of missing steps in O that are in B

- Order accuracy: $\frac{L}{K}$
  - *K* = Number of steps in common between O and B
  - *L* = Longest common subsequence between O and B

In some instances, time to select the correct course of action may be a relevant metric. The research team is considering including a Garden Path Scenario as part of our set of simulated scenario activities (Milltello et al. 2018). The Garden Path scenario injects evidence over time that, when aggregated, suggests a particular course of action. Initially, evidence for the correct course of action is minimal and exists among contradictory evidence. Over the course of the scenario, more evidence accumulates to send an increasingly strong signal for the correct course of action. This scenario evaluates operators' ability to break from an initial fixation on the incorrect course and adopt the new, correct course of action. If a Garden Path design is implemented, the research team will record Time to Correct Course of Action as our primary measure for this scenario.

Operator behaviors are an important objective class of measures to include in our assessment. However, although these measures tell us what operators did, it is important to note that they can only hint at the underlying cognition that might have motivated the behavior. Given this limitation, it is important to analyze these measures in concert with our cognitive assessments (critical thinking and SA).

5.0 References

Bergroth, Lasse, Harri Hakonen, and Timo Raita. "A survey of longest common subsequence algorithms." In *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000*, pp. 39-48. IEEE, 2000. https://doi.org/10.1109/SPIRE.2000.878178

"California Critical Thinking Disposition Inventory." Insight Assessment. Last modified 2025. https://insightassessment.com/iaproduct/california-critical-thinking-disposition-inventory/

"Cambridge International AS & A Level Thinking Skills 9694 Syllabus." Cambridge Assessment International Education. Last modified September, 2020. https://www.cambridgeinternational.org/Images/597412-2023-2025-syllabus.pdf

"Critical Thinking Skills Test." Insight Assessment. Last modified 2025. https://insightassessment.com/iaproduct/california-critical-thinking-skills-test/

Dunn, Dana S., Jane S. Halonen, and Randolph A. Smith, eds. *Teaching critical thinking in psychology: A handbook of best practices*. John Wiley & Sons, 2009. https://nibmehub.com/opac-service/pdf/read/Teaching%20Critical%20Thinking%20in%20Psychology%20_%20a%20handbook%20of%20best%20practices.pdf#page=85

Endsley, Mica R. "A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 63, no. 1 (2021): 124-150. https://doi.org/10.1177/0018720819875376

Endsley, Mica R. "Toward a theory of situation awareness in dynamic systems." In *Situational awareness*, pp. 9-42. Routledge, 2017.

Ennis, Robert, and Jason Millman. *Cornell Critical Thinking Test Level Z*. N.p.: Midwest Publications, 2005. https://www.criticalthinking.com/cornell-critical-thinking-test-level-z.html

Fero, Laura J., John M. O'Donnell, Thomas G. Zullo, Annette DeVito Dabbs, Julius Kitutu, Joseph T. Samosky, and Leslie A. Hoffman. "Critical thinking skills in nursing students: Comparison of simulation-based performance with metrics." *Journal of advanced nursing* 66, no. 10 (2010): 2182-2193. https://pmc.ncbi.nlm.nih.gov/articles/PMC5939573/pdf/nihms963923.pdf

Haimovitch, Liran. "Understandability: The Most Important Metric You're Not Tracking." InfoQ. Last modified June 6, 2020. https://www.infoq.com/articles/understandability-metric-not-tracking/

"The International Critical Thinking Essay Test." The Foundation for Critical Thinking. Last modified 2019. https://www.criticalthinking.org/pages/international-critical-thinking-test/619

Leighton, Jacqueline P. "Teaching and assessing deductive reasoning skills." *The Journal of experimental education* 74, no. 2 (2006): 107-136. https://doi.org/10.3200/JEXE.74.2.107-136

Militello, Laura G., Christen Sushereba, Simon Fernandez, David Bahner, and Emily S. Patterson. "A Systematic Method for Using Critical Decision Method Interviews to Generate a Garden Path Scenario for Human-in-the-Loop Simulations for Training and Evaluation of New Technologies." In *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 7, no. 1, pp. 69-74. Sage CA: Los Angeles, CA: SAGE Publications, 2018. https://doi.org/10.1177/2327857918071014

Newstead, Stephen E., Simon J. Handley, Clare Harley, Helen Wright, and Daniel Farrelly. "Individual differences in deductive reasoning." *The Quarterly Journal of Experimental Psychology Section A* 57, no. 1 (2004): 33-60. https://doi.org/10.1080/02724980343000116

"Shaping the Future Electric Grid." Pacific Northwest National Laboratory. https://www.pnnl.gov/sites/default/files/media/file/Shaping%20the%20Future%20Electric%20Grid.pdf

Speicher, Maximilian, and Johanna Jagow. "An Actionable And Reliable Usability Questionnaire With Only 7 Items: Inuit." Smashing Magazine. Last modified October 26, 2023. https://www.smashingmagazine.com/2023/10/actionable-reliable-usability-questionnaire-inuit/

"Test of Everyday Reasoning." Insight Assessment. Last modified 2026. https://insightassessment.com/iaproduct/test-of-everyday-reasoning/

"Watson-Glaser Critical Thinking Appraisal III (WG-III)." Pearson. Last modified April, 2020. https://www.talentlens.com/content/dam/school/global/Global-Talentlens/uk/CriticalThinking/CTHub/Pearson_Efficacy_Watson-Glaser_2020.pdf

Werner, Patrice Holden. 1991. "The Ennis-Weir Critical Thinking Essay Test: An Instrument for Testing and Teaching," *Journal of Reading* 34 (6): 494. https://www.proquest.com/scholarly-journals/ennis-weir-critical-thinking-essay-test/docview/216913834/se-2

Appendix A – Additional Critical Thinking Metrics

Here, we provide some additional critical thinking metrics that we feel should be considered in future assessments of critical thinking. For metrics associated with assessing the plausibility of a conclusion, more thorough investigation is needed to determine a feasible metric in a training environment. For metrics associated with forming a conclusion, we feel that these metrics, while important to critical thinking as a whole, are more focused on communication abilities than strictly critical thinking. We therefore feel that the other metrics listed in this document are better focused on critical thinking capabilities.

A.1 Additional Metrics for Analyzing a Conclusion or Action

A.1.1 Assessing the Plausibility of a Conclusion

In addition to applying methods of reasoning, the plausibility of the information or conclusion must be assessed through its credibility as well as whether any logical fallacies or cognitive biases may be at play.

Since credibility, logical fallacies, and cognitive biases are complex concepts with many different facets, more thorough investigation is needed to determine a metric that would be feasible to include in this training. Logical fallacies and cognitive biases will likely center on concepts like effort, intuition, association, priming, jumping to conclusions, and substitution.

A.1.1.1 Identifying Logical Fallacies and Cognitive Biases

Are there any logical fallacies or cognitive biases that are affecting the conclusion or proposed action?

Both logical fallacies and cognitive biases negatively affect a conclusion. Logical fallacies lead to erroneous or inappropriate connections within the conclusion, and cognitive biases lead to inappropriately weighting key information that support the conclusion or action. **Identifying such issues** in a conclusion or in support of an action, whether the originally proposed conclusion or action or another possible conclusion or action, helps determine which key information must at least be clarified. In some cases, those key information should be ignored altogether.

This metric aligns best with Segment 1 of the training course as the segment includes instruction on common pitfalls that adversely affect the decision-making process, which are related to logical fallacies and cognitive biases.

Potential Metric:

- Percent of correctly identified logical fallacies or cognitive biases from prompts:
$$\frac{number\ of\ logical\ fallacies\ or\ cognitive\ biases\ correctly\ identified}{total\ number\ of\ logical\ fallazies\ or\ cognitive\ biases\ in\ prompts} \times 100$$

A.1.1.2 Credibility

How credible is each key piece of information in support of the conclusion or action as well as the conclusion or action itself?

Credibility of a key piece of information, or a conclusion or action is determined by its truthfulness. If the truthfulness is unknown, as is often the case, then credibility extends from the provider of the information. How **trustworthy** are they, both in general and in the current context?

This metric aligns best with Segment 6 of the training course as the segment includes instruction on facts, secondhand information, and direct experiences as well as how to test assumptions related to each of these.

Potential Metrics:

- Trustworthiness assessment compared to known trustworthiness of information providers
- Completeness of verification of information from information providers

A.2 Metrics for Forming a Conclusion

With the originally provided conclusion or action as well as any proposed conclusions or actions analyzed, a final conclusion can be drawn or action determined. This conclusion or instructions for the action must be clear and focused to avoid any misunderstandings.

A.2.1 Stating a Conclusion

An effective conclusion or course of action is understandable and succinct.

A.2.1.1 Stating One's Point

Is the conclusion or course of action understandable to the intended audience?

The conclusion or course of action must be **understandable** to the intended audience, which means that it must be both **complete** and **clear**. Avoiding jargon that is not understood by the audience or asserting a conclusion that is perceived as incomplete will cause confusion and may even lead to the conclusion being missed entirely. For power grid operators, **stating their role and area of focus** helps ensure that the appropriate context is being given for the audience to build a complete understanding of the conclusion or course of action.

This metric aligns with Segment 7 of the training course as it focuses on making decisions and implementing a solution.

Suggested Metrics:

- Understandability of the point/conclusion (Haimovitch 2020; Speicher and Jagow 2023):
    - Completeness: Likert scale rating(s) (1 – 7) by the instructor(s) and/or other learner(s), with scores from multiple people averaged to generate an overall score if appropriate. Ratings will focus on components of the point/conclusion such as:
        - Who
        - When

        - Need
        - Risk
        - Criteria
    - Clarity: Likert scale rating (1 – 7) by the instructor(s) and/or other learner(s), with scores from multiple people averaged to generate an overall score if appropriate
- Role stated: True / False
- Area of focus stated: True / False

Additional Potential Metric:

- Organization of statements (Haimovitch 2020; Speicher and Jagow 2023): Likert scale rating (1 – 7) by the instructor(s) and/or other learner(s), with scores from multiple people averaged to generate an overall score if appropriate

A.2.1.2 Focus

How succinct is the conclusion?

Having **extraneous** information in the conclusion or course of action will make it harder for others to understand and may even distract from the conclusion or action. Minimizing extraneous information therefore aids in effectively communicating a conclusion or action.

This metric aligns with Segment 7 of the training course as it focuses on making decisions and implementing a solution.

Suggested Metric:

- Percent of information that was extraneous: $\frac{amount\ of\ extraneous\ information}{total\ amount\ of\ information} \times 100$

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov