Lesson Overview
This lesson explains how evaluators observe and collect relevant exercise data.

Lesson Objectives
After completing this lesson, you will be able to:

- Distinguish among the three types of reporting.
- Describe the three levels of performance analysis.
- List ways to minimize evaluator effects and errors.
- Distinguish between discussion- and operations-based exercise evaluation.
- Describe the purpose of the player hotwash.

This lesson should take approximately 50 minutes to complete.

Exercise Observation and Data Collection
Exercise observation and data collection is the second step of the exercise evaluation and improvement planning process.

This lesson divides Step 2 into four main topics:

1. The purpose and types of systematic observation.
2. Maximizing the effectiveness of observation.
3. Strategies for exercise evaluation.
4. Organizing data in preparation for analysis.

The Purpose of Systematic Observation
This portion of the lesson addresses the purpose and types of systematic observation. Systematic observation is the process of using standard forms such as Exercise Evaluation Guides (EEGs) to record performance of critical tasks.

Systematic observation ensures that data is consistent and well-organized. Such data is essential to the development of the After Action Report/Improvement Plan (AAR/IP), which identifies corrective actions to be implemented and tracked after the exercise.

Tracking progress on corrective actions is the cornerstone of the jurisdiction's Corrective Action Program (CAP). The CAP ensures that exercise findings are translated into concrete action, ensuring continual improvements to preparedness.

Types of Reporting
During an exercise, each evaluator performs three types of reporting:

1. **Descriptive Reporting.** Descriptive reporting is the direct observation and documentation of actions listed on evaluation forms. For example, consider a checklist item that asks whether the outgoing Operations Section Chief briefed his or her replacement. This item requires little subjective judgment on the part

of the evaluator. For that reason, it prompts descriptive reporting. Descriptive reporting typically yields reliable data.

2. **Inferential Reporting.** Inferential reporting requires an evaluator to arrive at a conclusion before recording information. For example, consider a checklist item that asks whether a capability is "adequate." In judging whether the capability is "adequate," the evaluator must first make an assumption about what "adequate" means. Since no two evaluators will make the exact same assumption, inferential reporting yields inconsistent data.
3. **Evaluative Reporting.** Evaluative reporting requires evaluators to assess performance on a scale of success. For example, consider an evaluation item that asks evaluators to rate the success of the incident commander's communications strategy. This item requires the evaluator to make an evaluative judgment. Reliable evaluative data is difficult to collect.

For the most part, evaluators will perform descriptive reporting.

Post-exercise activities ask the evaluator to assess data in relationship to exercise objectives. These assessments require inferential and evaluative judgments.

Observing Three Levels of Performance Analysis
During an exercise, evaluators collect data in multiple ways. They record their observations; collect data from records and logs; and attend the player hotwash and the Controller/Evaluator Debrief.

Whatever the data collection method, evaluators should perform three levels of performance analysis.

1. **Task-Level Analysis.** Task-level analysis focuses on specific, discrete actions. This analysis helps jurisdictions target plans, equipment, and training resources to improve performance on specific tasks. Tasks are often linked to performance measures designed to assist evaluators. For example, the "WMD/Hazardous Materials (HazMat) Response and Decontamination" capability contains the task "Implement mass decon operations." This task is accompanied by check boxes marked "Fully," "Partially," "Not," and "Not Applicable."

2. **Activity-Level Analysis.** Activities are groups of similar tasks that, when carried out according to plans and procedures, support a capability from the Target Capability List/Universal Task List (TCL/UTL). For example, the task "Implement mass decon operations" is part of the activity "Decontamination and Clean-Up/Recovery Operations." Other decontamination tasks also fall under this activity. Activity-level analysis focuses on whether all activities have been performed successfully and in accordance with plans, policies, procedures, and agreements.

Through this analysis, evaluators gain valuable insight into broad "themes" of successes or challenges in performing related tasks. Awareness of such themes is key to improving the performance of individual tasks, and thus demonstrating the associated capability.

3. **Capability-Level Analysis.** Capabilities are specific functionalities that support the high-level mission. They are combinations of elements such as personnel, planning, organization and leadership, equipment and systems, training, exercises, assessments, and corrective actions.

   When conducting capability-level analysis, evaluators examine whether performance on specific tasks and activities was sufficient to demonstrate the desired capability. Capability-level analysis is designed to assist managers and executives in developing operating plans and budgets, communicating with political officials, setting long range training and planning goals, and developing inter-agency and inter-jurisdictional agreements.

As you learned in Lesson 3, capabilities, activities, tasks, and performance measures are linked to an overall mission.

There are four homeland security missions: (1) preventing, (2) protecting against, (3) responding to, and (4) recovering from catastrophic events.

Note: The After Action Report/Improvement Plan (AAR/IP) focuses its analysis on activities and capabilities. It includes an analysis of tasks, however, to support root-cause analysis and recommendations for corrective action.

The Value of Coordination in Observation
This portion of the lesson explains techniques to maximize the effectiveness of observation.

In Lesson 3, you learned that training evaluators before the exercise enhances their ability to collect useful data. Coordinating evaluators during the exercise serves the same purpose.

Coordinating evaluators is especially important in large functional or full-scale exercises, when the exercise takes place at multiple locations.

Strong communication is the key to coordinating the efforts of evaluators. During exercises that last two or more shifts, outgoing evaluators should brief incoming evaluators. In addition, evaluators should have a standard procedure for communicating potential challenges to the Lead Evaluator.

Note: Coordination procedures should be developed before the exercise. They may be outlined in the Controller/Evaluator Handbook or communicated in the Controller/Evaluator briefing.

Avoiding Common Pitfalls of Evaluation
Evaluations are only effective if evaluators perform systematic observation and generate unbiased records. To ensure unbiased records, evaluators should avoid seven pitfalls of exercise evaluation:

- **Observer Drift.** Observer drift occurs when evaluators lose interest or a common frame of reference during an exercise. It is usually the result of fatigue or lack of motivation. Observer drift can be minimized by feedback from the Lead Evaluator, beverages and snacks, breaks, and rotational shifts of exercise observation.
- **Errors of Leniency.** Errors of leniency occur when evaluators have a tendency to rate all actions positively. It can be minimized by pre-exercise training.
- **Errors of Central Tendency.** Errors of central tendency occur when evaluators describe all activities as average in order to avoid making difficult decisions. It can be minimized by pre-exercise training.
- **Halo Effect.** The halo effect occurs when evaluators form a positive impression of a person or group early in the exercise and permit this impression to influence their observations. It can be minimized by pre-exercise training.
- **Hypercritical Effect.** The hypercritical effect occurs when evaluators believe it is their job to find something wrong, regardless of the players' performance. It can be minimized by pre-exercise training.
- **Contamination.** Contamination occurs when evaluators know how an activity was performed in earlier exercises and permit this knowledge to affect their expectations. It can be minimized by pre-exercise training.
- **Evaluator Bias.** Evaluator bias refers to errors that are traceable to characteristics of the evaluator. Evaluator bias can be minimized by careful selection of evaluators, or by employing multiple evaluators to observe the same functions.

Minimizing the Effect of Evaluators on Players
You just learned that data collection can be affected by an evaluator's prejudices. It can also be affected by his or her presence. It is well-documented that when evaluators observe exercises, the behavior of players may change. As a result, evaluators may observe atypical actions.

To reduce this effect, the evaluator should:

- Avoid recording any observations right away, so players become accustomed to the "intrusion."
- Arrive at the appropriate location before players do (this is particularly important if the exercise location is indoors instead of outdoors).

Evaluator presence may also influence players if the players anticipate what evaluators are looking for. For example, players may want to complete a task in a way that is not described in the emergency operations plan, but feel compelled to follow the planned procedure when the evaluator is watching.

Evaluators can minimize this impact by assuring players that the evaluation report will not reflect unfavorably on individuals.

When Evaluators Intervene in the Exercise
You just learned the importance of evaluators keeping a "low profile" during exercises. There are some occasions, however, when evaluators must draw attention to themselves. They may need to intervene to:

- Gather information that is unavailable elsewhere in order to accurately evaluate a capability.
- Clarify a situation that they did not understand.
- Prevent a potentially dangerous situation.

Whether evaluators may intervene during an exercise is decided when the exercise is designed. If evaluators may intervene, they should follow several guidelines:

- **Intervene only when necessary.** Remember that the evaluator's presence can distract players. With this in mind, the evaluator should:
    - Minimize questions by jotting them down and waiting to see if they are answered in the course of the exercise.
    - Ask questions at the player's convenience, such as during a lull in exercise play.
    - Ask essential questions quickly and let players return to their task.
- **Ask questions in language that the player understands.** If the objective or checklist item uses exercise jargon (for example, "When were players acclimated to the exercise?"), reword the question into simpler terms (for example, "When did the real action begin?").
- **Avoid leading questions.** A leading question is a question that prompts the responder to think one reply is better than another. For example, imagine that an evaluator is observing an evacuation. Consider the differences between the questions below:
    - **Leading Questions**
        - "Were the citizens evacuated within the appropriate timeframe?"
        - "Were the citizens evacuated quickly enough?"
    - **Non-Leading Questions**
        - "At what time did you begin evacuating the citizens?"
        - "At what time was the evacuation complete?"
- **Avoid prompting questions.** Prompting questions are leading questions that may affect overall exercise play. For example, consider the question, "Have you

begun evacuating the affected area yet?" This question may prompt players to act on your suggestion. A more objective question is "Are there any citizens at risk in the affected area?"

- **Avoid the role of advisor.** Under questioning, players may look to evaluators for guidance. To avoid the role of advisor, evaluators can ask:
    - "What would you do if I were not here?"
    - "What does the plan say?"
    - "What would you do in an actual occurrence?"

Tips for Reducing Evaluator Effects and Errors
In sum, basic guidelines for minimizing evaluator effects and errors include the following:

- Complete evaluator training before each exercise.
- Familiarize yourself with the various types of rating errors.
- Familiarize yourself with the evaluation checklists and report forms.
- If you are missing key data, ask the Lead Evaluator for help.
- Avoid making evaluative and inferential judgments during the exercise.
- Avoid conversations that could influence your impression of the exercise.
- Report obvious evaluator bias - in yourself or others - to the Lead Evaluator.

Use of Exercise Evaluation Guides (EEGS)
This portion of the lesson explains strategies for evaluating the exercise.

As you learned in Lesson 3, Exercise Evaluation Guides (EEGs) identify the activities, tasks, and performance measures that the evaluator should observe during the exercise.

Evaluators should complete the EEG so that:

- Events can be reconstructed at a later time (such as during summary sessions).
- Evaluators can conduct root-cause analyses of problems.

To ensure EEGs are fully complete, evaluators should:

- Synch their timekeeping with other evaluators before the exercise.
- Record the name and time of the exercise (as applicable).
- Log times accurately.
- Take notes on whether exercise simulations affect the observed task.

Complete EEGs are essential to the development of the After Action Report/Improvement Plan (AAR/IP).

Discussion-Based Exercise Evaluation
In a discussion-based exercise, evaluators should record discussions as they progress through the exercise. Because existing Exercise Evaluation Guides (EEGs) were

developed for operations-based exercises, they should be modified for use in discussion-based exercises.

While recording discussions, evaluators should pay special attention to:

- Issues identified by players.
- How players make decisions.
- Player roles and responsibilities.
- Player coordination and cooperation.
- Recommendations from the group.

Note: Both discussion-based and operations-based exercises are followed by a hotwash, but the content of the hotwash depends on the type of exercise. You will learn more about this distinction later in the lesson.

Operations-Based Exercise Evaluation
During an operations-based exercise, the main role of evaluators is to watch and record player actions.

While recording player actions, evaluators should pay special attention to:

- What actions took place.
- Who performed an action or made a decision.
- Where an action or decision took place.
- When an action or decision took place.
- Why an action was performed or a decision was made.
- How players performed an action or made a decision.

Note: The evaluator's role is to visually capture the actions of players. Evaluators may need to move around to watch events unfold from start to finish.

Tools for Collecting Data
This portion of the lesson will discuss strategies for organizing data in preparation for data analysis.

Evaluators should select evaluation tools that are appropriate to the exercise's location, size, length, and format. For example, videos, audio recordings, and photographs are particularly useful for gathering information in operations-based exercises. Exercise Evaluation Guides (EEGs) for discussion-based exercises should be customized accordingly.

Immediately after the exercise, evaluators should review their notes for gaps in information. If gaps exist, evaluators can fill them by gathering data from a variety of sources. EEGs are one source. Others include:

- Participant Feedback Forms.
- Player questionnaires.
- Event Logs.
- Telephone conversation records.
- Video or audio recordings.
- Copies of incoming, outgoing, and internal messages.
- Timelines.
- Evaluator folders, logs, and timeline.
- Player notes.
- Notes from controllers (in operations-based exercises).
- Incident reports.
- Notes from facilitators (in discussion-based exercises).
- Court recorders.
- Photographs.

Evaluators may also use lulls in the exercise to begin identifying and filling information gaps.

Player Hotwash
After the exercise, one or more player hotwash is held. It is attended by the Exercise Planning Team, players, evaluators, and facilitators or controllers. The player hotwash is an opportunity for players to describe their immediate impressions of demonstrated capabilities and the exercise itself. For this reason, it affords a valuable opportunity for evaluators to fill in gaps in their notes.

Player hotwashes allow time for players to address key topics, cross-disciplinary issues, or conflicting recommendations that were identified in earlier discussions. They are also an opportunity for players to comment on how well the exercise was planned and conducted.

Player hotwashes should be held as soon as possible after the exercise is complete, while player observations are still fresh. They are most effective when led by an experienced facilitator who can keep the discussion constructive and focused.

During the hotwash, evaluators, controllers and/or facilitators should distribute Participant Feedback Forms for players to submit.

For evaluators, a hotwash is an opportunity to collect player observations, clarify unclear points, and gather missing information. Although evaluators may be assigned to record a particular group discussion, they should capture information on cross-cutting issues.

A Controller/Evaluator Debrief is typically held after the player hotwash. Attended only by controllers, evaluators, and Exercise Planning Team members, the Controller/Evaluator Debrief is a forum to review and provide feedback on the exercise.

It should be a facilitated discussion that allows each person an opportunity to provide an overview of the functional area they observed, and to document both strengths and areas for improvement. Results of the Debrief should be captured for inclusion in the After Action Report (AAR).

Note: The content of the player hotwash depends on whether the exercise is discussion-based or operations-based.

- **Discussion-based.** In a discussion-based exercise, the player hotwash is typically held directly after the exercise. Attendees include the exercise players, the Exercise Planning Team, facilitators, and evaluators. The goal of this hotwash is to collect player observations on what occurred and why.

- **Operations-based.** In an operations-based exercise, the player hotwash is typically held on the last day of the exercise. Attendees include the exercise players, the Exercise Planning Team, controllers, and evaluators. The goal of this hotwash is to facilitate players in a self-assessment of the exercise play.

If a full-scale exercise has several sites, a hotwash should occur at each location.

Preliminary Analysis
After the exercise, each evaluator should gather his or her observations into key issues and a chronological narrative of events. When organized, these observations form the evaluator's preliminary analysis. Preliminary analyses feed the development of the After Action Report/Improvement Plan (AAR/IP).

At a minimum, an evaluator's preliminary analysis should include:

- A description of the assigned function or operation, analyzed by capability, activity, and critical task, as structured by the Exercise Evaluation Guides (EEGs).
- A documented record of significant evaluated actions (for example, an exercise event timeline).

Lesson Summary
In this lesson you learned:

Members of the Evaluation Team perform three types of reporting:

1. Descriptive Reporting.
2. Inferential reporting.
3. Evaluative reporting.

In an exercise evaluation, evaluators collect data for the three levels of performance analysis:

1. Capability-level performance.
2. Activity-level performance.
3. Task-level performance.

Key pitfalls that evaluators should avoid include:

- Observer drift.
- Errors of leniency.
- Errors of central tendency.
- The halo effect.
- The hypercritical effect.
- Contamination.
- Evaluator bias.

Many of these pitfalls can be mitigated by effective pre-exercise training.

The purpose of the hotwash is to collect player feedback and address key topics, cross-disciplinary issues, or conflicting recommendations that were identified in earlier discussions.

While collecting data in a discussion-based exercise, evaluators should pay special attention to:

- Issues identified by players.
- How players make decisions.
- Player roles and responsibilities.
- Player coordination and cooperation.
- Recommendations from the group.

While collecting data in an operations-based exercise, evaluators should pay special attention to what, who, where, when, why, and how actions and decisions took place.

Immediately after the exercise, evaluators should create preliminary analyses of their findings.