Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions

Recommendations for Developing an Improvement and Evaluation Process

LESLIE ADRIENNE PAYNE, ANDREA M. ABLER, SUSAN G. STRAUS, JASON MICHEL ETCHEGARAY
About This Report

This report documents research and analysis conducted as part of a project entitled *Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions*, sponsored by the Integration Branch of the Federal Emergency Management Agency (FEMA) within the Workforce Development Division of the Field Operations Directorate. The findings should be of interest to FEMA, given that the purpose of the project was to identify gaps and challenges in assessing the current state of the FEMA Qualification System (FQS) and provide recommendations that will allow FEMA to evaluate and improve FQS.

This research was sponsored by FEMA’s Integration Branch and conducted in the Disaster Management and Resilience Program of the RAND Homeland Security Research Division (HSRD), which operates the Homeland Security Operational Analysis Center (HSOAC).

About the Homeland Security Operational Analysis Center

The Homeland Security Act of 2002 (Public Law 107-296, § 305, as codified at 6 U.S.C. § 185) authorizes the Secretary of Homeland Security, acting through the Under Secretary for Science and Technology, to establish one or more federally funded research and development centers (FFRDCs) to provide independent analysis of homeland security issues. The RAND Corporation operates HSOAC as an FFRDC for the U.S. Department of Homeland Security (DHS) under contract 70RSAT22D00000001.

The HSOAC FFRDC provides the government with independent and objective analyses and advice in core areas important to the department in support of policy development, decisionmaking, alternative approaches, and new ideas on issues of significance. HSOAC also works with and supports other federal, state, local, tribal, and public- and private-sector organizations that make up the homeland security enterprise. HSOAC’s research is undertaken by mutual consent with DHS and organized as a set of discrete tasks. This report presents the results of research and analysis conducted under 70FA4022F00000193, FEMA Qualification System. The results presented in this report do not necessarily reflect official DHS opinion or policy.

For more information on HSRD, see www.rand.org/hsrd.

For more information on this publication, see www.rand.org/t/RRA2460-1.
Acknowledgments

We thank our FEMA Integration Branch sponsors for their guidance and encouragement throughout this effort. This includes Tara Bailey, Garrett Coley-Freeman, Stephanie DeLorenzo, Cheraye Ellis, Clare Anne Fagan, Ivy Fry, Dedrick Jackson, Sr., Christopher McGough, Timothy Robinson, Emma Simon, Patrick Smith, and Melissa Ventresca. We also extend our appreciation to the many FEMA incident workforce staff who participated in interviews.

At HSOAC, we thank Jessie Riposo, Shelly Culbertson, Jason Thomas Barnosky, and Evania Baginski for their assistance and support throughout this effort. We also thank Isabelle Winston for her commitment to our data collection effort, along with the internal and external reviewers, Christopher Nelson and Bill Carwile, for their valuable feedback on earlier versions of this report.
Summary

To fulfill its mission objectives, the Federal Emergency Management Agency (FEMA) requires a well-trained workforce. The Post-Katrina Emergency Management Reform Act of 2006 mandated that FEMA develop a personnel credentialing process for incident workforce (IW) personnel responding to disaster situations.¹ FEMA, in turn, created the FEMA Qualification System (FQS), a credentialing process that focuses on standardizing personnel abilities through training and applied experience.²

Following concerns raised by the Government Accountability Office (GAO) about the efficiency and effectiveness of FQS to provide the needed structure to qualify individuals for their assigned positions,³ FEMA’s Integration Branch asked the Homeland Security Operational Analysis Center (HSOAC) to identify existing FQS gaps and areas for improvement and examine ways to measure and monitor the effectiveness of FQS in the future. This report details HSOAC’s approach to addressing both issues and the related findings so that FEMA can ensure the development of employees who provide vital disaster recovery services to the United States and its people.

Approach

Our HSOAC study team used a largely qualitative approach to data collection and conducted expert elicitation interviews with 83 IW subject-matter experts (SMEs) between August 2022 and February 2023. Additionally, we reviewed literature, documents, and government-furnished information (GFI) obtained from FEMA’s Integration Branch and those we interviewed.

This study was composed of three research tasks: Tasks 1 and 2 were designed to understand what is working well and what needs improvement across five core areas of FQS: Position Task Books (PTBs), Coaches and Evaluators (C&Es), training, the Qualification Review Board (QRB), and the Deployment Tracking System (DTS). Task 3 was focused on whether and how effectiveness is measured, assessed, or evaluated across these areas. We report strengths and areas for improvement that SMEs identified, propose metrics to assess the health of FQS metrics, and offer other courses of action (CoAs) that can improve aspects of the FQS process.

¹ Senate Homeland Security and Governmental Affairs Committee, 2006.
² FQS is a program of four sequential steps: Process Entry, Position Task Book (PTB) Requirements, Qualification Process, and Possible Progression (FEMA, 2022a). These steps are supported by “sections” that help IW personnel transition from being candidates for a position to earning qualification to perform the position.
Findings

Findings in the form of strengths, areas for improvement, and approaches to measuring the effectiveness of FQS are provided throughout this report. For the latter, interviewees largely described the absence of consistent and codified efforts to measure many of the gaps we identified.

Given this lack of rigorous metrics, it is not possible to create a fully informed judgment on the current health of FQS that is backed by supporting data. Although we offer recommendations to help FEMA move in the direction of eventually assessing the health of FQS, the lack of current metrics and data render any ability to currently assess FQS incomplete at best. Despite these challenges, we learned over the course of the research that there is a genuine interest in improving FQS, which speaks to an improvement-oriented culture that can be leveraged in the future to measure the effectiveness of and subsequently improve FQS.

Position Task Books

Strengths

As shared in interviews, many in the IW feel that the PTB is constructive for laying out the tasks required for given positions. PTBs were also described as being constructive for professional growth and career progression because they help FEMA IW personnel track their progress in a systematic way. A limited number of interviewees highlighted the process by which SMEs and other experts are consulted to increase efficiency with the revision process as a strength.

Areas for Improvement

SME responses and our review of PTBs and other GFI point to areas for improvement. One SME described an erosion of trust in PTB qualification because of feeling that the system is no longer reliable.

A review of a sample of PTBs indicated that not all sections contain complete information, including those that should always be complete, such as the number of required endorsements. Having blank sections calls into question the overall quality of the PTB and does not provide clear guidance to the trainee/candidate (T/C) and the C&Es about what is needed for task completion.

Although indicators are optional in the PTB rubric, when they were present in PTBs, they were not always written as behavioral statements, as would be consistent with their stated purpose in the FEMA Qualification System Handbook.4

One key theme raised by SMEs was that some cadres and C&Es feel pressured to have high PTB completion rates for reports to leadership and Congress.

4 FEMA, 2022b.
Another common theme from SMEs was the need to improve the PTB revision process, though some SMEs explained that earlier and bumpier approaches to revisions have improved somewhat. Others offered less-positive views of the process, citing the need to increase standardization and develop a smoother process.

The need for alignment between the content of PTBs and training is critical, and there were varied perspectives about how well this is being done. Some interviewees indicated that training is aligned well with PTBs, and others were unsure whether alignment is occurring.

Measuring the Quality of Position Task Books

According to SMEs, there are some tools, such as templates and glossaries, used at the headquarters (HQ) level to assess PTB completeness during the revision process, but SMEs identified no other tools being used to systematically measure PTB effectiveness. Table S.1 proposes measures and related actions to evaluate the health of PTBs.

Recommendations for other CoAs to improve PTBs include the following:

- Conduct surveys or focus groups to understand reasons for difficulties in completing PTBs.
- Develop and codify expectations of ownership for SMEs, cadre staff, and HQ staff working on creating new PTBs or reviewing and revising existing PTBs.
- Disseminate a stand-alone graphic to the IW showing the interdependencies between the trainer, C&E, and responder.
- Use multiple forms of communication (e.g., official FEMA memoranda, administrator and senior leader email messages, training curricula) to emphasize the need to maintain good communication and integrity in C&E sign-offs for the PTB process to be effective and efficient.
- Develop a formalized process to incorporate minor but critical PTB changes that arise between official revision cycles.

<table>
<thead>
<tr>
<th>TABLE S.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recommended Measures and Actions to Assess the Health of Position Task Books</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gap</th>
<th>Recommendations on What to Measure or Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTB quality</td>
<td>• Continue using tools to improve PTB effectiveness (e.g., specific, measurable, achievable, relevant, and time-bound [SMART] goals rubric, task template, glossary) during PTB creation and revision</td>
</tr>
<tr>
<td></td>
<td>• Develop a standardized rubric or checklist consisting of required PTB elements to ensure accuracy and completeness</td>
</tr>
<tr>
<td>Quantification of PTB-training alignment</td>
<td>• Develop a standardized rubric or checklist that documents when tasks from PTBs and content from training are aligned</td>
</tr>
</tbody>
</table>
Coaches and Evaluators

Strengths
As described in FEMA Directive 010-11 and the FEMA Coach and Evaluator Guide, a strength of the C&E program is that it is designed to be vital to the training efforts of IW personnel and serves as a validation mechanism for FQS to ensure that those T/Cs qualified for positions are able to perform tasks fundamental to those positions. Interviewees reported that C&Es bring significant experience to the evaluation process.

Areas for Improvement
GAO’s May 2020 report FEMA Disaster Workforce: Actions Needed to Address Deployment and Staff Development Challenges cited FEMA data showing that “at the start of deployments during the 2017 and 2018 disaster seasons, 36 percent of staff did not have an official assigned to coach and evaluate task performance—the primary mechanism the agency depends on for coaching.” There were several areas where interviewees validated points raised by this GAO report. Some examples of similar concerns include the following:

- C&Es appear to sometimes sign off on PTB tasks in an inconsistent and selective manner, although one interviewee cited the subjective nature of some PTB tasks.
- C&Es are overburdened by a multitude of deployment-related tasks, the presence of which might affect their effectiveness as C&Es.
- There is a lack of qualified C&Es in the field during deployments.
- The absence of evaluative processes for C&Es directly contributes to problems such as bias and a lack of accountability.
- One cadre described their method of deploying dedicated C&Es (i.e., C&Es with no deployment-related duties outside their C&E responsibilities) that allows PTBs to be signed off more expeditiously during deployments. Similarly, the 2020 GAO report describes the success of a dedicated C&E program piloted in 2019.

Interviewees also raised concerns about the content and frequency of training for C&Es. These were described primarily as resource issues.

Measuring the Effectiveness of Coaches and Evaluators
SME responses indicate that FEMA does not systematically assess C&E effectiveness. Table S.2 proposes measures and related actions to assess the health of C&Es.

---

5 FEMA, 2019b; FEMA, 2019a.
6 GAO, 2020, cover page summary.
7 GAO, 2020, cover page summary.
TABLE S.2
Recommended Measures and Actions to Assess the Health of the Coaches and Evaluators Program

<table>
<thead>
<tr>
<th>Gap</th>
<th>Recommendations on What to Measure or Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficient number of C&amp;Es</td>
<td>• Determine appropriate staffing ratios for the IW as a whole and for incident types; analyze administrative data to determine whether staffing ratios are achieved</td>
</tr>
<tr>
<td></td>
<td>• Analyze potential gaps between qualification status (i.e., successful completion of C&amp;E Qualification Course as recorded in DTS) and role assignments for shortages of C&amp;Es</td>
</tr>
<tr>
<td>Quality of C&amp;E ratings</td>
<td>• Analyze C&amp;E ratings for common rating errors and for possible differences in ratings by protected class or other factors</td>
</tr>
<tr>
<td></td>
<td>• Analyze interrater reliability—consistency among different C&amp;Es’ ratings of the same candidates</td>
</tr>
<tr>
<td></td>
<td>• Require supervisor oversight and performance review of C&amp;Es</td>
</tr>
<tr>
<td></td>
<td>• Implement 360-degree feedback for developmental, not administrative, purposes</td>
</tr>
<tr>
<td></td>
<td>• Develop clear definitions of proficiency in checklists to evaluate T/C proficiency on PTBs</td>
</tr>
<tr>
<td>Inconsistent, selective, or biased C&amp;E signoff</td>
<td></td>
</tr>
<tr>
<td>C&amp;E sign off before T/C is proficient</td>
<td></td>
</tr>
<tr>
<td>Lack of evaluation of C&amp;Es</td>
<td></td>
</tr>
<tr>
<td>Subjective nature of PTB tasks</td>
<td></td>
</tr>
</tbody>
</table>

Recommendations for other CoAs to improve C&E processes include the following:

- Use multiple forms of communication (e.g., official FEMA memoranda, administrator and senior leader email messages, training curricula) to publicize clear expectations for C&Es and their role in maintaining IW readiness.
- Create a mechanism for C&Es provide feedback to training personnel at the FEMA Incident Workforce Academy (FIWA) regarding T/C readiness.
- Hire more C&Es and expand C&E training to focus on obscure PTB tasks and performing evaluations of T/Cs under challenging circumstances.
- If none exists, establish a grievance/redress process for trainees who perceive unfair treatment or evaluations from C&Es.
- Deploy individuals with a dedicated C&E role, removing deployment-related duties outside their C&E responsibilities.

Training
Strengths
FEMA’s FQS branch uses a systematic curriculum mapping process to ensure that training aligns with PTBs. One interviewee mentioned that changes to training during the coronavirus disease 2019 pandemic to spread out courses gave students more time to absorb what they have learned.
Areas for Improvement
Several SMEs noted the tension between FEMA's need to deploy IW personnel and those IW personnel potentially being undertrained to perform key duties. Some also commented on the lack of refresher training and lack of opportunities to deploy, resulting in skill atrophy. There are concerns that there are not enough instructors to staff the courses, courses are not offered frequently, and cadres might prioritize access to their own members over those from other cadres.

Other challenges we heard related to collaboration on training development, delivery, and evaluation, and we found it difficult to determine responsibility for different aspects of training. SMEs also raised concerns about training evaluation. Finally, several interviewees discussed the fragmented nature of information technology systems to support training; relatedly, some pointed to a need for a learning management system or learning content management system (LCMS).

Measuring the Effectiveness of Training Processes and Outcomes
Interviewees indicated that there are some processes for evaluating the effectiveness of training, but these are not always comprehensive. For example, although FIWA conducts a needs analysis to determine annual training schedules, results are suboptimal because they receive only limited feedback from cadres. Regarding the Kirkpatrick levels of evaluation, we learned about a lack of centralized collection or analysis of Level 1 data, nascent attempts to conduct Level 2 assessments, and an absence of Level 3 and Level 4 assessments.

Table S.3 proposes measures and related actions to assess the health of training. Recommendations for other CoAs to improve PTBs include the following:

- Analyze the quality of knowledge tests and provide training, training aids, and access to experts to assist test developers in improving test quality.
- Establish policy that delineates ownership of various aspects of training, including development, delivery, and evaluation.
- Investigate options, such as individual, simulation-based training, to maintain skill proficiency between deployments.
- Adopt an enterprisewide LCMS that maximizes needs as identified by FEMA.
- Seek fiscal resources for sufficient FEMA and/or contractor personnel to increase course availability and support use of metrics.

Qualification Review Boards
Strengths
Interviewees recognize the intent of the QRB to provide independent evaluations of candidates, outside the cadres. Some SMEs also discussed how the QRB serves as a checkpoint to ensure that a candidate meets all of the required benchmarks, for example, by providing rec-
ommendations for further training or additional deployments and by prompting a thorough review and preparation of the candidate by the cadre.

In addition, some SMEs describe recent changes to the QRB that strengthen its utility. Another SME reported that there is a lot of effort being put into updating the QRB process for it to be useful for the cadres.

Areas for Improvement

Many SMEs expressed the sentiment that the QRB adds little value to the qualification process because, regardless of the QRB’s recommendation, the cadre Certifying Authority (CA) holds final decision authority on whether a candidate is qualified. Other issues that SMEs raised included the following:

- There is a lack of sufficient feedback to improve a candidate’s qualification packet (though this statement seems to be in direct conflict with a SME who reported that the QRB provides a “full list of findings and a report” to the cadres). It is unclear why this discrepancy exists and where it originates.
- QRB members do not have the correct expertise to judge the proficiency of cadre candidates.
- The QRB process is too time consuming.
- There are inconsistencies in official FEMA guidance. (For example, the conflicting guidance on whether deployments are required for candidate positions in the FQS Guide that cadres follow and the QRB Handbook.)
- QRB recommendations can be biased.

TABLE S.3

<table>
<thead>
<tr>
<th>Gap</th>
<th>Recommended Measures and Actions to Assess the Health of Training</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recommendations on What to Measure or Evaluate</td>
</tr>
<tr>
<td>Availability of classes</td>
<td>• Continue needs analysis and identify strategies for increased cadre involvement</td>
</tr>
<tr>
<td>Standard and systematic Level 1 evaluations</td>
<td>• Create a combination of standardized questions and questions that cadres can customize that are administered online</td>
</tr>
<tr>
<td></td>
<td>• Administer online assessments of pre-training and post-training self-efficacy for task performance</td>
</tr>
<tr>
<td>Quality of Level 2 assessments</td>
<td>• Conduct practical exercises (work samples) that reflect the job requirement to assess behavioral outcomes during training, using well-defined checklists and clear options for rating proficiency</td>
</tr>
<tr>
<td></td>
<td>• Create carefully constructed written tests of substantive knowledge for course concepts</td>
</tr>
<tr>
<td></td>
<td>• Create pre-tests and post-tests to assess improvement in learning</td>
</tr>
<tr>
<td></td>
<td>• Analyze the psychometric properties of tests, ideally using item response theory</td>
</tr>
<tr>
<td></td>
<td>• Retain the full range of test scores, not only “pass/fail” outcomes, to assess training effectiveness</td>
</tr>
</tbody>
</table>
Measuring the Quality of QRB Decisions

Discussions with SMEs indicate that there are no measures to gauge QRB effectiveness. Table S.4 proposes measures and related actions to assess the health of QRB processes.

Recommendations for other CoAs to improve QRB processes include the following:

- Analyze return on investment of the QRB, considering perceived utility, costs of member and candidate time, impartiality of recommendations, timeliness, and potential conflicts with other FEMA goals.
- Consider an alternate model that better addresses the technical nuances of each cadre and job position.
- Continue implementing changes, such as updating documentation and providing additional services to cadres, to enhance cadres’ investment in the process.
- Update the QRB Handbook to reflect policy regarding deployment requirements and qualification in the FQS Guide.

Deployment Tracking System

Strengths
A SME familiar with querying DTS commented on the ability of the system to contain and extract valuable information.

Areas for Improvement
We learned from interviews that DTS contains a lot of data, but the quality and meaningfulness are uncertain. SMEs also noted difficulties with structuring queries to answer questions of interest, with some noting that the system can provide a wealth of information and others unsure how to effectively use the system.

Some interviewees reported that they rely on the decision made by the algorithm in DTS in terms of selecting IW personnel for deployment, whereas other SMEs frequently request IW personnel by name rather than trusting the algorithm. Some reasons for submitting by-

### TABLE S.4
**Recommended Measures and Actions to Assess the Health of Qualification Review Boards Processes**

<table>
<thead>
<tr>
<th>Gap</th>
<th>Recommendations on What to Measure or Evaluate</th>
</tr>
</thead>
</table>
| Lack of trust in value and legitimacy of QRB recommendations | • Conduct regular, brief online surveys of stakeholders’ perceptions of QRB decision processes and outcomes and communicate results to stakeholders  
• Analyze differences in QRB decisions/outcomes through the lens of protected class or other factors  
• Consider using a quantitative model that combines numeric ratings on inputs currently used by the QRB to make qualification recommendations |
name requests included knowing the capability of and trusting specific IW personnel and the need to help those closer to PTB completion achieve additional experience.

Measuring the Effectiveness of the Deployment Tracking System

The gaps or issues that SMEs raised point largely to potential needs for training and system redesign. However, one gap indicates the need to monitor users’ perceptions of DTS usefulness. Discussions with SMEs did not reveal measures that FEMA uses to assess DTS effectiveness. Table S.5 proposes measures and related actions to assess the health of DTS.

Recommendations for other CoAs to improve DTS include the following:

- Enlist assistance from the contractor that developed DTS to create an online training aid that explains the DTS algorithm and offers best practices for usage.
- Modify DTS to provide more-specific information about the percentage of task completion and/or which PTBs personnel have completed.

Summary and Conclusions

Overall, FQS has a fair number of strengths, and stakeholders see value in the concept behind FQS and its component parts. However, in practice, the component pieces do not always function as designed. In addition, we found very few systematic measures and evaluation activities to gauge whether processes in FQS components are implemented effectively and in ways that align with FEMA’s goals. We have proposed a variety of measures to assess the health of the five components and other CoAs to improve gaps in FQS. In addition, in the concluding chapter of this report, we rate recommendations for both metrics and other CoAs in terms of two factors: a temporal dimension (time needed for implementation) and an operational factor (ease of implementation [e.g., in terms of required resources and cost]). We rated the recommendations as less or more for each factor. For example, in the CoAs to improve training, we consider the recommendation to provide training assistance to test developers as requiring less time and being less difficult to implement, whereas we consider adoption of an enterprisewide LCMS as requiring more time and being more difficult to implement.

<table>
<thead>
<tr>
<th>TABLE S.5</th>
<th>Recommended Measures and Actions to Assess the Health of the Deployment Tracking System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap</td>
<td>Recommendations on What to Measure or Evaluate</td>
</tr>
<tr>
<td>Satisfaction with DTS</td>
<td>• Conduct regular online surveys of users’ perceptions of DTS effectiveness and communicate results to stakeholders</td>
</tr>
<tr>
<td>Trust of the DTS algorithm</td>
<td>• Determine frequency with which users select IW personnel for deployment by name rather than relying on the DTS algorithm and conduct a qualitative analysis of justifications</td>
</tr>
</tbody>
</table>
However, ultimately, we acknowledge that FEMA is in the best position to determine which recommendations to pursue and in what time frame. Thus, FEMA can use the results from our categorization effort as a starting point for their own discussions about how to prioritize the recommendations and the ease or hurdles they might encounter in efforts to implement them.
Contents

About This Report iii
Summary v
Approach v
Findings vi
Summary and Conclusions xiii

CHAPTER 1
Introduction 1
Purpose of This Report 1
Overview of the FEMA Qualification System 2
Methodological Approach 3
Deciding on Dedoose 6
Project Scope 8
Project Limitations 8
Organization of This Report 8

CHAPTER 2
Position Task Books 11
Strengths of Position Task Books 11
Areas for Improvement 13
Assessing the Effectiveness of Position Task Books: Findings and Recommendations on What to Measure or Evaluate 17
Course of Action Recommendations 18

CHAPTER 3
Coach and Evaluator Program 21
Strengths of Coaches and Evaluators 22
Areas for Improvement 22
Assessing the Effectiveness of Coaches and Evaluators: Findings and Recommendations on What to Measure or Evaluate 26
Course of Action Recommendations 28

CHAPTER 4
Training 31
Strengths of Training 32
Areas for Improvement 32
Assessing the Effectiveness of Training: Findings and Recommendations on What to Measure or Evaluate 37
Course of Action Recommendations 44
Figures and Tables

Figures

1.1. FEMA Cadres .................................................................................. 2
1.2. Research Questions Addressed by HSOAC ................................... 4
1.3. Literature and Documents Reviewed by HSOAC .......................... 5
1.4. Round 1 Themes ........................................................................... 7
1.5. Round 2 Themes ........................................................................... 8
4.1. Kirkpatrick Levels of Training Evaluation .................................... 37
5.1. Role Played by the Qualification Review Board in the Trainee/Candidate Qualification Process ................................................................. 45
7.1. Measurement and Evaluation Recommendations by Time and Ease of Implementation ................................................................. 59
7.2. Course of Action Recommendations by Time and Ease of Implementation ...... 62

Tables

S.1. Recommended Measures and Actions to Assess the Health of Position Task Books...................................................................................... vii
S.2. Recommended Measures and Actions to Assess the Health of the Coaches and Evaluators Program ................................................................. ix
S.3. Recommended Measures and Actions to Assess the Health of Training ...... xi
S.4. Recommended Measures and Actions to Assess the Health of Qualification Review Boards Processes ............................................................. xii
S.5. Recommended Measures and Actions to Assess the Health of the Deployment Tracking System ................................................................. xiii
1.1. Summary of Round 1 Data Statistics, Interviews Held August–October 2022 .... 7
1.2. Summary of Round 2 Data Statistics, Interviews Held November 2022–February 2023 ................................................................. 7
2.1. Recommended Measures and Actions to Assess the Health of Position Task Books...................................................................................... 18
3.1. Recommended Measures and Actions to Assess the Health of the Coaches and Evaluators Program ................................................................. 27
4.1. Recommended Measures and Actions to Assess the Health of Training ...... 42
4.2. Sample Item from Lytell et al., 2017 .................................................... 42
5.1. Recommended Measures and Actions to Assess the Health of Qualification Review Board Processes ............................................................. 50
6.1. Recommended Measures and Actions to Assess the Health of the Deployment Tracking System ................................................................. 54
7.1. Full List of Findings per FEMA Qualification System Area...................... 56
7.2. Full List of Measurement and Evaluation Recommendations .................... 59
7.3. Full List of Course of Action Recommendations ........................................ 63
A.1. Dedoose Codebook for Round 2 Interviews ............................................. 70
CHAPTER 1

Introduction

The Federal Emergency Management Agency (FEMA) is responsible for “helping people before, during, and after disasters,” and fundamental to this mission is the need for a well-trained workforce to execute mission objectives. The Post-Katrina Emergency Management Reform Act of 2006 mandated that FEMA develop a personnel credentialing process for incident workforce (IW) personnel responding to disaster situations. FEMA, in turn, created the FEMA Qualification System (FQS) to serve as this credentialing process and focused on standardizing personnel abilities through training and applied experience. FQS represents a potentially holistic solution for performance development so that a capable and ready IW can be mobilized to respond to disasters.

Purpose of This Report

In its 2020 report FEMA Disaster Workforce: Actions Needed to Address Deployment and Staff Development Challenges, the Government Accountability Office (GAO) noted that FEMA was working toward—with a suspense date of March 31, 2021—a plan to communicate employee knowledge, skills, and abilities to field leaders and managers, with the ability to communicate being partially dependent on a robust FQS process. The report also identified challenges with and concerns about the efficiency and effectiveness of FQS to provide the needed structure to qualify individuals for their assigned positions. FEMA’s Integration Branch asked the Homeland Security Operational Analysis Center (H5OAC), a federally funded research and development center (FFRDC) operated by the RAND Corporation, to identify current gaps and areas for improvement in FQS and examine ways that FQS can be evaluated so that it can be improved. This report details H5OAC’s approach to addressing both issues—current areas for improvement in FQS and ways to assess it in the future—so that FEMA can make adjustments when supporting the development of those employees providing vital disaster recovery

1 FEMA, 2023.
2 Senate Homeland Security and Governmental Affairs Committee, 2006.
3 FEMA, 2022a.
services to our country and its people. FEMA often works in tandem with state, local, tribal, and territorial responders and others when providing these critical services to affected populations and communities. However, the assessments conducted over the course of the study, in addition to the recommendations offered, apply solely to FEMA's IW.

Overview of the FEMA Qualification System

FQS consists of a series of interdependent steps designed to standardize the capabilities of the IW through experience, training, and demonstrated performance. FQS consists of four specific steps that FEMA IW members progress through in the FQS program: Process Entry, Position Task Book (PTB) Requirements, Qualification Process, and Possible Progression. These steps are supported by sections—which refer to processes or systems—that help FEMA IW personnel transition from being candidates for a position to earning qualification to perform the position. Sections that provide the foundational elements of FQS include PTBs, Coaches and Evaluators (C&Es), training, Qualification Review Board (QRB), and the Deployment Tracking System (DTS).

IW personnel are assigned to one of 23 cadres (see Figure 1.1) that differ based on functional responsibilities and mission to support FEMA in meeting the National Preparedness

---

5 FEMA, 2022a.
Goal.6 Once cadres justify the need for personnel via a force structure requirement that is approved by FEMA leadership, IWs are onboarded. The IW employee is not considered a candidate for the position until they either become qualified for the position or are no longer pursuing the position.7 The candidate’s pursuit of the position is formally documented by a PTB, which contains the tasks one needs to successfully perform in a position (additional elements of PTBs are discussed in Chapter 2). Candidates learn how to perform tasks via training, which is developed and delivered either by the Training Section at the FEMA Incident Workforce Academy (FIWA) or by the cadres to which they are assigned, who collaborate with the Training Development Section (TDS) of the FQS Branch.

Candidates also learn how to perform tasks through on-the-job experience, with guidance from cadre experts during disaster deployments who serve in the role of C&Es. C&Es coach candidates to help them understand how to improve performance and evaluate candidates to document that they are sufficiently skilled in a task for the position. Once candidates for Tier 1 (e.g., Federal Coordinating Officers, Section Chiefs) and Tier 2 (e.g., group supervisors, unit leaders) positions have completed a PTB, the PTB is validated, and then a multi-step process is followed whereby multiple experts review the candidate to ensure proficiency in the position. Steps in the process include review by a member from the cadre, followed by a Certifying Official, and then by a QRB. The QRB’s review of the candidate results in a recommendation about candidate qualification to the CA, who then documents their qualification decision in DTS.

Methodological Approach

FEMA’s Integration Branch asked HSOAC to help it understand current areas for improvement in FQS and the best ways to assess the program in the future. To address this aim, our HSOAC study team used a largely qualitative approach to data collection and conducted expert elicitation interviews with IW subject-matter experts (SMEs) between August 2022 and February 2023.8 These interviews were designed to help HSOAC answer the research questions delineated in the study’s Technical Execution Plan and endorsed by FEMA (see Figure 1.2). In total, we spoke with 83 individuals.

Additionally, we reviewed literature, documents, and government-furnished information (GFI) obtained from FEMA’s Integration Branch and interview participants. We leveraged information from GFI when preparing for and conducting SME interviews and integrated

---

6 “A secure and resilient nation with the capabilities required across the whole community to prevent, protect against, mitigate, respond to, and recover from the threats and hazards that pose the greatest risk” (FEMA, 2019c, p. 8).
7 FEMA, 2022a.
8 Our study team defines SMEs to be individuals with deep and direct experience in, or knowledge about, the five FQS areas around which our research revolved.
Expert Interviews

In addition to reviewing the documents shown in Figure 1.3, we conducted two rounds of expert elicitation interviews to address the study’s key research questions. For Round 1, which focused on Tasks 1 and 2 spanning August–October 2022, we conducted 28 interviews and spoke with a total of 65 SMEs across various FEMA areas (e.g., cadres, FIWA, other headquarters [HQ] branches) to identify areas for improvement in the FQS program and develop future-oriented recommendations to guide improvement. Although some of these interviews had a single participant (n = 11), others were group conversations with multiple participants (n = 17). Interview participants were chosen with intent based on the breadth and depth of information needed about the FQS program. This purposeful sampling technique entailed first envisioning the category of FEMA personnel we needed to engage (e.g., personnel at training locations, cadre members, personnel in specific branches), then working with Integration Branch personnel to schedule interviews with individuals from each category.9

For all interviews, we used a semi-structured interview approach examining the pressing gaps in FQS and its sections, ways that FQS is continuously monitored, and additional factors that need to be considered when improving FQS in the future. At the end of each interview, we asked SMEs for recommendations of other SMEs we should consider interviewing. These interviews were generally 60 minutes in length, conducted via Microsoft Teams, and SMEs provided verbal consent for participation in the study prior to participating in the interview. Interviews were typically led by one researcher while another took notes. The second

9 Purposeful sampling is often conducted in contrast to snowball sampling, a common qualitative method whereby one respondent recommends another, and another, until the sample snowballs to a large number of respondents and a data saturation level is reached.
researcher took near-verbatim notes to capture the SMEs’ responses.\footnote{To rapidly analyze this set of interviews and develop initial findings to quickly share with the Integration Branch, we spent time after the interviews discussing what we learned and developing themes that reflected the findings from the interviews. We shared these preliminary findings with the Integration Branch three months into the project via a briefing.} As described in detail in the appendix, we analyzed interview data using a mixed methods (qualitative and quantitative) coding and analysis software platform called Dedoose. Dedoose allowed us to review, collate, identify, and analyze key themes that cut across these 28 conversations.

Our Round 2 interviews focused on Task 3 objectives and spanned November 2022–February 2023. During this period, we conducted 18 interviews with FEMA personnel and spoke with a total of 31 SMEs. As in Round 1, we used a purposeful sampling technique to
ensure that we spoke with individuals from a wide variety of cadres and HQ offices. Also like Round 1, some Round 2 interviews had a single participant ($n = 10$), while others were group conversations with multiple participants ($n = 8$). Of these 18 interviews, eight were with distinct cadres and ten were with HQ personnel. We also conducted two semi-structured, individual interviews with participants from organizations external to FEMA: the United States Coast Guard and the Civil Air Patrol. These interviews focused on current ways to measure the quality and/or effectiveness of five FQS processes: PTBs, C&Es, training, QRB, and DTS.

Deciding on Dedoose

At the outset of our research planning process, we reached unanimous consensus on the utility of qualitative data analysis, or coding software. Those benefits include but are not limited to (1) quickly and more accurately interpreting statements, (2) assigning codes to words and phrases to capture the content of the response, (3) sorting and analyzing a large number of responses, and (4) graphically depicting thematic relationships among key results by both quantitative and qualitative means.

We used Dedoose for analyzing Round 1 and Round 2 interviews because of its accessibility, affordability, analytical capabilities, and usability. Founded and supported by a team of professionally trained and active social-science researchers, Dedoose is a web-based application for analyzing content in various types of media. See the appendix for the Dedoose codebooks we developed and used when analyzing interview comments, in addition to other components of our coding strategy, such as the use of inter-rater reliability (IRR), which ensures that researchers share a common understanding of codes and apply these codes to text in a consistent and accurate manner. Throughout the report, we lightly edited quotes when necessary to improve clarity and readability.

Table 1.1 summarizes key statistics from Dedoose for Round 1 interviews. Round 1 interview themes were grouped in seven nonexclusive topics (i.e., a single SME response can be associated with more than one of these topics): evaluation, monitoring, and metrics; inherent FQS frictions; fragmented application of FQS; PTB completion difficulties; PTB and training content; C&Es; and lack of resources and personnel. Figure 1.4 reports the frequency with which each of these topics was mentioned in Round 1 interviews.

Table 1.2 summarizes key statistics for Round 2 interviews. In Round 2, interview themes were grouped into the following non-exclusive topics: training; PTBs; evaluation and assessment; C&Es; purpose, description, and process; weaknesses; and strengths. Figure 1.5 reports the frequency with which each of these topics was mentioned in Round 2 interviews.

---

TABLE 1.1
Summary of Round 1 Data Statistics, Interviews Held August–October 2022

<table>
<thead>
<tr>
<th>Measure</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of interviews</td>
<td>28</td>
</tr>
<tr>
<td>Total number of SMEs engaged</td>
<td>65</td>
</tr>
<tr>
<td>Total number of Dedoose codes (or themes) we looked for in</td>
<td>45</td>
</tr>
<tr>
<td>interview transcripts</td>
<td></td>
</tr>
<tr>
<td>Total number of excerpts created in Dedoose</td>
<td>365</td>
</tr>
<tr>
<td>Total number of codes applied</td>
<td>986</td>
</tr>
</tbody>
</table>

FIGURE 1.4
Round 1 Themes

<table>
<thead>
<tr>
<th>Times mentioned by respondents</th>
<th>Evaluation, monitoring, and metrics</th>
<th>Inherent frictions</th>
<th>Fragmented application</th>
<th>PTB (completion difficulties)</th>
<th>PTB (training content)</th>
<th>C&amp;Es</th>
<th>Lack of resources and personnel</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>73</td>
<td>64</td>
<td>54</td>
<td>57</td>
<td>48</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

TABLE 1.2
Summary of Round 2 Data Statistics, Interviews Held November 2022–February 2023

<table>
<thead>
<tr>
<th>Measure</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of interviews</td>
<td>18</td>
</tr>
<tr>
<td>Total number of SMEs engaged</td>
<td>31</td>
</tr>
<tr>
<td>Total number of Dedoose codes (or themes) we looked for in interview transcripts</td>
<td>24</td>
</tr>
<tr>
<td>Total number of excerpts created in Dedoose</td>
<td>365</td>
</tr>
<tr>
<td>Total number of codes applied</td>
<td>1,086</td>
</tr>
</tbody>
</table>
Project Scope

For project scoping purposes, we sought to understand a wide variety of gaps and areas for improvement that FEMA SMEs were willing to share during Round 1 interviews. We then iterated with the Integration Branch and agreed on focusing the Round 2 interviews on SMEs’ perceptions of and methods to assess the effectiveness of the five elements of FQS—PTBs, C&Es, training, QRB, and DTS—where effectiveness can entail quality of processes or outcomes and throughput.

Project Limitations

There were a variety of limitations—most of which related to access to interviewees—that affected aspects of this research study. These are discussed in detail in the “Conclusions” section of this report.

Organization of This Report

We organized the report in six additional chapters that detail our findings, recommendations, and conclusions. We structured Chapters 2 through 6 to focus on various aspects of five FQS sections—PTBs, C&Es, training, QRB, and DTS, respectively. In each, we present findings in three categories: strengths, areas for improvement (i.e., gaps), and assessing effec-
tiveness (the extent to which it is or is not measured). Each of these chapters concludes with recommendations for measurement and other courses of action (CoAs) to improve each FQS section.

How We Present Findings
Our strategy of categorizing the findings by strengths, areas for improvement, and assessment of effectiveness relates to how we structured this research.

As described in Figure 1.2, Tasks 1 and 2 of this research were designed to understand what is working well in each of the five areas of FQS and what needs improvement. Interview questions in support of these two tasks focused on providing HSOAC a robust picture of how FQS is functioning for the IW. During these Round 1 interviews, some respondents organically offered comments on the extent to which effectiveness was assessed in the PTB process, in relation to C&Es, and in relation to training (with few comments provided on QRB and DTS). However, most comments centered on strengths and weaknesses. Our analysis shows that participants offered disproportionately more comments on weaknesses compared with strengths.

Round 2 interviews were in support of Task 3 of this study, which focused on whether and how effectiveness is measured, assessed, or evaluated across FQS areas. We posed highly tailored questions to interviewees to understand the presence or absence of metrics used to evaluate the PTB process, C&Es, training, QRB, and DTS and what that means for the IW. The dominant theme encountered was that effectiveness across FQS needs to be measured, but currently, for the most part, it is not. Perhaps it was because of a lack of measures to assess effectiveness that conversations veered back to the topics of strengths and weaknesses.

Given the above, the portion of each chapter that discusses effectiveness, or Task 3 findings, (1) includes a summary statement about the data found, (2) offers findings in the form of strengths and areas for improvement, and (3) includes suggestions from HSOAC in the form of a measures table for how FEMA could measure effectiveness going forward.

Regarding the measures table, it is possible that there are existing measures that we did not identify in documentation that was not accessible, not provided to the team, or not discussed in SME interviews. However, in the absence of those measures, we present what we view to be viable ways for FEMA to assess the effectiveness of these five FQS sections going forward. We view our proposed measures as the start of a conversation with FEMA about how to assess and improve the health of FQS, with additional input from FEMA needed as it further expands this list.

Chapter 7 collates our findings and recommendations. Three tables are provided for easy reference. Table 7.1 summarizes our findings on FQS strengths and weaknesses, Table 7.2 summarizes our recommendations on measuring effectiveness across FQS, and Table 7.3 presents CoA recommendations that relate to each of the five FQS areas. The recommendations presented in Table 7.2 pertain to suggested approaches to measuring FQS effectiveness, whereas the recommendations in Table 7.3 pertain to other CoAs that FEMA can implement.
to address other needs for improvement that SMEs identified in FQS elements. Chapter 7 also assesses both sets of recommendations in terms of timeliness and ease of implementation via an implementation matrix or *quad chart*. Last, Chapter 7 presents areas for improvement that cut across several FQS areas with suggestions for remediation and our concluding thoughts on the research and findings. The appendix offers additional details about our approach to conducting and analyzing SME interviews, including our codebook.
Position Task Books

A PTB serves as a progress tracker for candidates as they complete requirements to be qualified in a position. The PTB is accessed electronically, and should define

- tasks that need to be completed
- competencies that align with the task
- behavioral examples of the task
- complexity codes that identify the environments in which the tasks must be demonstrated by the candidate, ranging from disaster levels (D-1 to D-3) to steady state
- the number of C&E endorsements needed to verify that the candidate can perform the task
- indicators that provide contextual information to help candidates understand how to perform the task and help evaluators assess performance
- whether the task is considered a core competency that exists across multiple PTBs.

Strengths of Position Task Books

As we describe in this chapter and those that follow, each of the five FQS areas we assessed has processes or elements that are working well (strengths) and gaps or areas for improvement. Our Round 2 analysis revealed that interviewees described more strengths related to PTBs than strengths related to C&Es, training, the QRB, or DTS. This equates to a preponderance of positive sentiment about PTBs compared with the other topics.

Individual and group interviews with eight different cadres resulted in the similar perspective that many in the IW say that the PTB is constructive for laying out the tasks required for given positions. As an example, a SME noted, “the strength, in theory, is that PTBs help us identify the key points of the tasks or the key points of the positions that an individual needs to learn to be able to be proficient at the particular job.”

---

1 Round 2 group interview (#10). The number found in parentheses after each interview citation relates to the interview participant. For example, Round 2 group interview (#10) refers to the 10th interview engagement during Round 2, in which more than one SME participated.
Another SME offered a similar statement: “You're 100 percent right that everything is based around the position task book. The position task book is the guideline for getting qualified and for what the qualified should be able to complete.” A different SME from a group interview shared a similar view but added an admonition: “The idea of using task books to lay out expectations of a position is good. Our execution of those task books is lacking.”

SMEs also described PTBs as constructive for professional growth and career progression because they provide a foundation for FEMA IW personnel to track their progress in a systematic way. An interviewee explained that the PTB can help someone see how “to go from a specialist through [the necessary] intermediary steps to branch director; I think that’s a positive thing.” A SME with management experience relatedly remarked that the PTB allows “us to communicate to our workforce what is expected of them and what we expect them to be able to do to perform their job functions and progress in our hierarchy and structure.”

We also heard a unique perspective from one SME about the PTB’s contributions to promoting equity and fairness: “For me, the task books are the foundation of equity. [The task book] ensures that we are providing a single standard across our entire workforce for what we say someone’s skills and abilities are that they have to achieve and maintain.”

Several interviewees pointed out observable strengths of the PTB revision process, such as the process by which SMEs and other experts are consulted to increase efficacy. One SME from a focus group interview described assembling a “coalition of experts” to help with revisions: “Our cadre facilitates subject-matter expert panels—SMEs that are qualified to help us review training and PTBs—to validate whether or not a PTB is successful.” Relatedly, the Integration Branch shared with us that an external contractor is reviewing all tasks across PTBs to determine which ones appear to be well written and should be retained. Another SME described an ongoing effort to improve alignment of PTBs and training:

So, we do have an opportunity to review PTBs especially when we’re about to update or edit an existing course that’s tied to that PTB, or if we’re creating a brand-new course. We’ve been working towards it but as of yet we do not have one course associated with every single PTB that we possess. The very first step we take before even starting to work on a course development or revamp is to look at the PTB first because the course will then directly be mapped to those PTB tasks.

---
2 Round 2 group interview (#6).
3 Round 2 group interview (#9).
4 Round 2 SME interview (#15).
5 Round 2 SME interview (#10).
6 Round 2 SME interview (#5).
7 Round 2 SME interview (#5).
8 Round 2 group interview (#10).
Areas for Improvement

SME responses and our review of PTBs and other GFI point to areas for improvement, including comprehensiveness of PTBs, PTB completeness, PTB revision processes, and alignment of PTBs with training. We learned that some IW personnel experience stress and anxiety about PTB completion. In at least four conversations spanning our Round 1 and Round 2 interviews, SMEs described the strain of getting PTB tasks signed off or demonstrating proficiency on select tasks amid skeptical reviewers. For example, a Round 2 SME described responders as having anxiety about the PTB and suggested that this might be because of unfamiliarity with the process.9 Similarly, a Round 1 focus group participant described that an erosion of trust in PTB qualification creates stress for responders: “Some of us feel that the system is not reliable anymore. I think it does put additional stressors on the people that really are truly qualified. They almost kind of have to prove themselves when they go out because many of us are so unsure of the system.”10

Comprehensiveness of Position Task Books

FEMA provided us with a sample of PTBs for us to better understand their content. In our review, we noticed that not all sections in the PTB contained complete information. However, the FEMA Qualification System Handbook noted that some omissions are to be expected (for example, competencies are not identified for newer positions, and indicators are optional or additional guidance to help C&Es assess the task).11 Other sections that should always be complete, such as number of endorsements, were sometimes blank. Having blank sections calls into question the overall quality of the PTB and does not provide clear guidance to the trainee/candidate (T/C) or the C&E about what is needed for task completion. Although indicators are optional, when they were present in PTBs, they were written as (1) behavioral statements (which seems to be consistent with their stated purpose in the FEMA Qualification System Handbook12), (2) phrases or terms, or (3) yes or no questions. An indicator that is listed as a yes or no question does not provide actual guidance to the T/C or C&E. These three approaches to documenting indicators lack consistency in communicating information to T/Cs and C&E about task expectations, and the latter two are inconsistent with the guidance in the handbook.

9 Round 2 SME interview (#4).
10 Round 1 group interview (#16).
11 FEMA, 2022b.
12 FEMA, 2022b.
Position Task Book Completeness

One key theme we identified in SMEs’ comments is that some cadres and C&Es feel pressured to have high PTB completion rates for reports to leadership and Congress. In Chapter 3, we explore this in more detail and discuss the consequences of such pressures in terms of C&Es giving inflated ratings or endorsing PTB completion before T/Cs are proficient. However, incomplete PTBs do not always indicate nonproficiency. There are a variety of factors that might result in incomplete PTBs even when responders are qualified (e.g., there might be a lack of C&Es available in the field to sign off on PTBs, periodic PTB revisions might make completing a PTB a moving target, a T/C might be deemed proficient by their cadre but held up by technicalities during the QRB process, or new hires might have proficiency from previous experience that is not reflected in PTBs). There are also reasons it is difficult to reach proficiency, which could include a lack of training opportunities, a lack of disasters at the incidence levels required for proficiency (e.g., deployed to D-3 but need D-1 experience to sign off), or the deployment of T/Cs to positions other than open PTBs (they could receive no credit for tasks in PTB because of another deployed position).

Position Task Book Revision Processes

Another common theme from SMEs was the need to improve the PTB revision process. However, we heard various opinions about how worrisome this problem is, with some SMEs explaining that earlier and bumpier approaches to revisions have improved somewhat. A group interview participant described how the timeliness of the revision process is not perfect, but it is better than it was previously: “I believe the current process in place is timely because cadres are able to review their PTBs essentially yearly. Because we are starting our iterative updates with our PTBs, I think things will be a lot smoother in the future.” Others offered less-positive view of the process, citing the need to increase standardization and develop a smoother process. Interviewees often lamented what they viewed as an ad hoc revision process, though some credited select cadres with trying to transition to more formal processes. For example, a SME explaining the lack of standardization said,

Over the last few years, we’ve made minor adjustments to PTBs, but that was ad hoc, and now we’re trying to look at each position as a whole in advance of developing or updating to ensure the tasks in the task book align with what the major things are that the role needs to accomplish. We do that through SMEs, we solicit feedback from different folks,

---

13 D-3 is a disaster requiring a moderate amount of federal assistance, while D-1 is a disaster requiring an extreme amount of federal assistance (FEMA, undated).

14 For example, one SME explained that their cadre spends a lot of time “updating every one of our PTBs and associating them with clear standards” (Round 2 interview [#1]).

15 Round 2 group interview (#11).
we have working groups with people who have field experience, using them to evaluate the activities that those positions do in the field and need to be reflected in the PTB.\textsuperscript{16}

Another SME remarked on the inefficiency of updating PTBs; this participant suggested that incorporating DTS in the PTB revision process would enhance efficiency, particularly when there are many revisions to make:

Whenever we get a big task to revise or delete, I would like to upload or export it to DTS and have DTS do it for me, so I don’t have to do it line by line by line. I feel like that would have more potential.\textsuperscript{17}

Last, two other categories of PTB problems echo issues mentioned elsewhere in this report, indicating that some areas for improvement are needed across multiple FQS processes. First, some interview participants cited challenges collaborating with others to revise PTBs, despite their best intentions to increase efficiency and ease, for example:

The challenge is they don’t always work in the field all the time so they may have a HQ perspective of what the task should be. Sometimes there’s a misalignment with what people in the field may say they do or should do.\textsuperscript{18}

Second, interviewees described a need for sufficient staff levels to appropriately update and revise PTBs. It appears that this gap, though perhaps more problematic in the past, is potentially less so as of the time of our interviews. A Round 2 SME said,

When we started FQS, in [our] cadre it was just a couple of people that were responsible for the entire cadre and PTBs and C&E, and training, and everything, and we didn’t have the infrastructure to start the program off right. What we’re doing now is trying to fix our implementation of FQS by providing more staff [who] can update the PTBs to reflect changes in position requirements.\textsuperscript{19}

### Aligning Position Task Books and Training

Given the purpose of PTBs, tight alignment between the content of PTBs and training is critical. We included specific questions about alignment in our interview protocols, including questions about how alignment is achieved, documented, and assessed. We obtained mixed results, with some interview participants indicating that training is aligned with PTBs and

\textsuperscript{16} Round 2 SME interview (#9).
\textsuperscript{17} Round 2 SME interview (#15).
\textsuperscript{18} Round 2 SME interview (#10).
\textsuperscript{19} Round 2 group interview (#9).
others who were unsure about whether alignment is occurring. Those who discussed alignment between training and PTBs commented,

When we develop training, the trainings are aligned with tasks . . . . When we’re developing a plan of instruction and have different modules/units, we have to align the content of those modules with a task in the PTB, so that’s one way that we ensure the content is driven by the PTB.\(^\text{20}\)

A different SME noted that the possibility of misalignment existed:

Here’s the potential for the training not to match the content of the PTB, and even if that matches there’s no guarantee that the training itself is teaching people what they’re actually supposed to be learning. So, it feels like there [are] several areas where there can be a disconnect.\(^\text{21}\)

When we probed about specific ways that SMEs captured alignment between training and PTBs, the responses were more focused on process and less on quantifying the extent to which alignment occurred, with responses ranging from general activities (e.g., “listing which tasks are being addressed by each learning activity by each module”)\(^\text{22}\) to specific ones (e.g., “FQS Branch asks our cadre to validate these tasks [and] make sure that everything in there is accurate. We’ll go back to our SMEs and take a look at those tasks”\(^\text{23}\) before sending confirmation that the tasks are correct). As one SME noted, “Everything that’s in the PTB should be reflected in training.”\(^\text{24}\) It was difficult for us to discern ways that those overseeing the PTB-training alignment process quantified and evaluated whether this was truly occurring. As a further example of a process, a different Round 2 SME stated that the FQS Branch ensures that cadres also have training and development (TD) personnel present on joint calls to discuss changes to the PTB: “When those meetings are scheduled, the goal is to make sure there’s someone from TDs on that call. So, they’re able to speak the training language.”\(^\text{25}\) Furthermore, the process used to align training with tasks was described as conducting a “[curriculum] mapping of the course to the task. There could be one or more tasks per objective or more than one objective per task, but the goal is to align training and tasks listed in the PTBs.”\(^\text{26}\)

\(^{20}\) Round 2 group interview (#9).
\(^{21}\) Round 1 group interview (#8).
\(^{22}\) Round 2 group interview (#12).
\(^{23}\) Round 2 group interview (#5).
\(^{24}\) Round 2 group interview (#8).
\(^{25}\) Round 2 SME interview (#15).
\(^{26}\) Round 2 group interview (#20).
Despite the presence of the curriculum mapping process described above, which we note as a strength in Chapter 4, we would be remiss if we did not also note discrepancies we heard concerning a foundational element of PTBs (i.e., the tasks). We learned that a “task is defined by the cadre. Each cadre has different task language that they use,” which implies variability in how cadres conceptualize a task.\textsuperscript{27} Complicating matters is that cadres might not have the expertise to write effective task statements. As one SME stated, “FQS defers to the cadre to determine what’s in the PTB, but cadres are not experts in doing any kind of study or knowing how to write a task that is objective . . . so cadres look to FQS to make sure those are clear and not duplicative.”\textsuperscript{28}

Assessing the Effectiveness of Position Task Books: Findings and Recommendations on What to Measure or Evaluate

Interviews with SMEs indicated that there are currently some tools used to improve the effectiveness of PTBs, mainly implemented at the HQ level during the PTB revision process. These include the use of templates and glossaries and ensuring tasks line up with specific, measurable, achievable, relevant, and time-bound (SMART) goals.\textsuperscript{29} Apart from the use of these tools, there was little indication from our interviews of systematic measures of PTB effectiveness. Many cadres discussed their use of SMEs in the PTB revision process to ensure alignment between PTBs and field requirements or pointed to their mapping processes during PTB revision to ensure alignment between PTBs and training curricula.\textsuperscript{30} However, the interviewees were not aware of any specific metrics used to evaluate effectiveness. One SME described it in this way: “In FEMA there is really no underpinning to the position task book. We manage tasks as a cadre but there hasn’t been a systematic approach to being able to clearly articulate for each task how it’s evaluated.”\textsuperscript{31} Another stated succinctly, “we do not do any evaluation of PTB effectiveness.”\textsuperscript{32}

In Table 2.1, we recommend additional measures and actions that FEMA can use to evaluate the effectiveness of PTBs.

Regarding PTB quality, we propose that FEMA create dedicated PTB development and revision teams with a standardized checklist to evaluate the accuracy and completeness of elements of the PTBs, along with guidance and examples showing correct and incorrect construction or composition of the elements. Finally, a process and accompanying checklist by

\textsuperscript{27} Round 1 SME interview (#2).
\textsuperscript{28} Round 2 group interview (#9).
\textsuperscript{29} Round 2 SME interview (#15).
\textsuperscript{30} Round 2 SME interviews (#1, #6, #9, #10).
\textsuperscript{31} Round 2 SME interview (#1).
\textsuperscript{32} Round 2 SME interview (#9).
which one can document the link between tasks from PTBs and content from training is needed for mapping and transparency purposes.

### Course of Action Recommendations

The recommendations below were informed by the interviews and relate to CoAs that FEMA can adopt to improve PTBs. They differ from those related to measurement and evaluation. Recommendations are as follows:

1. Collect insights on the variety of factors that lead to incomplete and complete PTBs by either creating a short survey to distribute among IW personnel or convening focus groups. Publicize the results of these efforts via official FEMA memoranda and commit to mitigation measures.

2. Develop and codify a standard set of expectations of ownership for SMEs, cadre staff, and HQ staff working on creating new PTBs or reviewing and revising existing PTBs, to ensure that the individuals involved understand their roles and how they relate to the roles and responsibilities of others.

3. Create and internally circulate through the IW a stand-alone graphic showing the interdependencies among the trainer, C&E, and responder.

4. Through official FEMA memoranda, official emails from the administrator, IW-wide emails from senior leaders, and training curricula, emphasize the need to maintain good communication and integrity in C&E sign-offs for the PTB process to be effective and efficient.\(^{33}\)

---

\(^{33}\) Round 2 interviews (#5, #15).
5. Develop and implement a formalized method to process minor but critical PTB changes that arise between official PTB revision cycles.\textsuperscript{34}

\textsuperscript{34} A Round 2 interviewee (#5) offered,

It would be great if there was some mechanism for some of the more minor changes. They may be minor but important. If it’s just the changing out of, you know, a specific form but having a formalized process for having a PTB changed currently, we’ll send an e-mail and say, “hey, you know, this is change that we’d like to see,” but I think having a formalized process that can be used kind of intermittently. That would be helpful because there are some things that can wait for a larger process to begin. But there’s others that really need to be a bit more immediate.”
CHAPTER 3

Coach and Evaluator Program

The C&E program is focused on three overarching principles: (1) contribute to the development of a qualified IW that assists disaster survivors and executes FEMA’s mission, (2) create a process focused on consistency and fairness that allows IW to qualify for positions, and (3) assist in the development of standardized qualifications for IW personnel.1 IW personnel in cadres, who have either been deployed in a position (for at least 30 days) or been qualified for a position (for at least six months), are eligible to apply to become a C&E via cadre management, who in turn provide nominations to the C&E program manager in the WDD, which is ultimately responsible for administering the C&E program.2 Once selected to be a C&E, IW personnel will complete initial C&E training provided by WDD that is supplemented with at least three hours of additional refresher training every two years. Once trained, cadres in concert with WDD ensure that C&Es are deployed to disasters to provide coaching and evaluation to IW personnel with open PTBs.3 During deployment, C&Es are expected to serve as IW personnel in their qualified position to provide disaster relief and serve as a C&E to help T/Cs progress through their PTBs.4

The C&E program is recognized as a critical component of FQS. As one interviewee remarked, “[a] huge part of FQS is C&E practices in the field. It is arguably the biggest part of FQS.”5 Although this was representative of several perspectives we encountered, our conversations with IW individuals, groups, and focus group participants suggest a nuanced picture regarding C&Es and the T/Cs they assess. As described below, although some spoke favorably about aspects of the C&E program and its personnel, others delineated challenges that require mitigation for the program to reach its full potential and continue serving the qualification process.

---

1 FEMA, 2019b.
2 FEMA, 2019b.
3 FEMA, 2019b.
4 We discuss T/Cs together, similar to the way they are discussed in the FEMA Qualification System Guide (FEMA, 2022a), though there are differences between the two. Trainee refers to an employee seeking qualification who does not have current qualification for a position. Candidate refers to an employee who is qualified for a FQS position and is seeking qualification in another position.
5 Round 1 SME interview (#3).
Strengths of Coaches and Evaluators

As described in FEMA Directive 010-11 and the FEMA Coach and Evaluator Guide, a strength of the C&E program is that it is designed to be vital to the training efforts of IW personnel—via the hands-on coaching and evaluating that is provided—and serves as a validation mechanism for FQS to ensure that those T/Cs qualified for positions are able to perform fundamental tasks.

Interviewees described a wealth of experience brought by C&Es to the evaluation process, with some drawing clear distinctions between mentors in the field and the higher level of fidelity produced by C&Es. One SME offered,

Those who are our coach evaluators [are] truly the ones who have gone above and beyond, and we feel can not only teach [and] mentor but can really say that somebody else has mastered the task. So, they are true experts in the field who have been doing it for rather a long time.7

This individual further stated,

We have pretty strict guidance if you want to become a C&E. You have to be qualified in your position for at least a year and actually deployed over 120 days in that qualified position as a base to be a C&E . . . . There’s a difference between our coach evaluators and our mentors.8

A similar sentiment offered by another Round 2 interviewee was that

one of the single most important aspects of that qualification system is the evaluator. So, we engage with our coach and evaluators on a consistent basis. We are very selective in who and how individuals become coach[es] and evaluators.9

Areas for Improvement

Both interview participants and findings from GAO’s May 2020 report FEMA Disaster Workforce: Actions Needed to Address Deployment and Staff Development Challenges point to areas

---

6 FEMA, 2019b; FEMA, 2019a.
7 Round 2 SME interview (#6).
8 Round 2 SME interview (#6). Noted the variance between the interviewee’s description of qualification requirements and what is found in FEMA’s C&E Directive, which states that qualified C&Es must either deploy in a position for 30 days or be qualified in a position for at least six months.
9 Round 2 group interview (#5).
for improvement in aspects of the C&E program.\textsuperscript{10} We present two general areas for improvement: number of C&Es and accuracy of C&E ratings. In addition to discussing the consequences of these gaps, we identify possible root causes of these issues, such as pressures affecting the organizational culture, social pressures resulting from a model in which C&Es sign off on each other’s PTBs, ambiguity in PTB tasks, and insufficient training and evaluation of C&Es.

**Number of Coaches and Evaluators**

SMEs cited problems related to the dearth of qualified C&Es in the field during deployments. As one interviewee stated, they regularly hear from the IW that “there was no one . . . willing to sign my PTB.”\textsuperscript{11} Another interviewee remarked, “I think, to be candid, the problem has been a matter of resources . . . . Most of what I hear is some concern that there [are not] enough coach evaluators to sign off on open task books, especially the larger Type One kind of disasters.”\textsuperscript{12} One cadre reported it has difficulty training and qualifying T/Cs, noting, “I only have two [C&Es] right now who are evaluating T/Cs.”\textsuperscript{13} Another SME described a shortage of C&Es capable of assessing senior leader positions,\textsuperscript{14}

> There are senior leadership positions . . . [and] it’s really difficult to get coach evaluators for those positions. I used to [have a management position] and struggled greatly with trying to have leadership out in the field that would devote the time to it. Some flat out would refuse to coach evaluate . . . . We’re struggling trying to get our senior leaders.

Another described a lack of C&Es in the field despite consistently training and qualifying individuals to be C&Es:

> We train and qualify people to be C&Es all of the time. We generate C&Es but there are not enough. I can’t speak to how we staff a disaster, but do we really want [a] capability? And if you do then [you] need a lot of coaching and [to] figure out cadence of that capability.\textsuperscript{15}

SMEs’ comments about the need for personnel are consistent with findings in the 2020 GAO report, which cited FEMA data showing, “at the start of deployments during the 2017

\textsuperscript{10} GAO, 2020.

\textsuperscript{11} Round 2 group interview (#8).

\textsuperscript{12} Round 1 SME interview (#9).

\textsuperscript{13} Round 2 SME interview (#6).

\textsuperscript{14} Round 2 group interview (#10).

\textsuperscript{15} Round 1 SME interview (#3).
and 2018 disaster seasons, 36 percent of staff did not have an official assigned to coach and evaluate task performance—the primary mechanism the agency depends on for coaching.\textsuperscript{16}

Having enough C&Es to properly evaluate in disaster situations is challenging given the cadence and operational issues inherent in those situations, as one SME noted:

Some of most glaring [challenges are] being able to get C&E staff to the field to be able to sign off on PTBs when we’re in the thralls of a disaster. C&Es are not necessarily requested in the beginning, and it’s difficult to get support staff out there where C&E is the only thing they’re doing. If it’s a big disaster season, it’s pretty much chaos getting everyone where they need to go to get things set up and running. It’s pretty much all hands on deck. It’s critical information, you need to know how to do it, but it’s hard to evaluate and coach people because of the time constraint. You’re busy making sure you’re completing the mission. It’s hard to evaluate a new person doing that when we’re under the gun to get something up and running. Mentoring, teaching, it’s very hard to get that done right off the bat.\textsuperscript{17}

The shortage of C&Es has multiple effects on the IW, including long lags in PTB approvals and C&E being overburdened by a multitude of deployment-related tasks that detract from their availability to perform C&E duties. To combat these challenges, one cadre reported using a “strike team model” that deploys C&Es who “are not responsible for any disaster work, only coaching and evaluating.” These C&Es “slowly make their way through the entire disaster to ensure everyone has an opportunity to be coached and evaluated on the work they’re doing.” According to this cadre, this model “allows us to get the [PTBs] signed a little bit faster.”\textsuperscript{18} This echoes information published in the 2020 GAO report, which describes a pilot exercise conducted by the National Disaster Recovery Support (NDRS) cadre. In 2019, NDRS deployed C&Es “solely in that position,” with an understanding from cadre management that the individual “was not to be used for other disaster-related responsibilities.”\textsuperscript{19} According to GAO, FEMA described the pilot as a success because “the coach and evaluator was able to devote time to proper training and answering any questions presented.”\textsuperscript{20}

**Accuracy of Coach and Evaluator Position Task Book Ratings**

Both the GAO report and our interviews pointed to concerns about the accuracy of PTB ratings and completion rates—specifically, that C&Es appear to sometimes inflate PTB rat-

\textsuperscript{16} GAO, 2020, p. 1, cover page summary.

\textsuperscript{17} Round 2 group interview (#10).

\textsuperscript{18} Round 2 group interview (#8).

\textsuperscript{19} GAO, 2020.

\textsuperscript{20} GAO, 2020, p. 39.
Coaching and Evaluating Program

ings or sign off on PTB tasks in an inconsistent or selective (rather than objective) manner. Interviewees described instances of PTB tasks that were signed off on or endorsed by C&Es when the T/C failed to demonstrate proficiency on those tasks. This becomes apparent when other C&Es later encounter responders in the field who appear unable to complete tasks independently, despite having PTBs that indicate otherwise.

Our interviews offered several possible explanations for inaccurate and inconsistent evaluations. First, as mentioned in Chapter 2, interviewees noted that cadres and C&Es feel pressured to approve PTBs to ensure high completion rates. This results in a pencil-whipping or check-the-box culture. Additionally, the low number of C&Es in the field and resulting burdens on them might lead to rushed evaluations. One person indicated that, because C&Es evaluate each other, they might be uncomfortable providing corrective feedback given that someone who is acting as a C&E in one situation might be evaluated by the very person they are evaluating in a subsequent situation. Inconsistency in ratings might also occur because different C&Es have different standards for performance. As a Round 1 SME stated:

How do you ensure consistency where one C&E reviews the task [and says] “I’m not signing off on that”? I don’t know how good we are with evaluating the C&Es to say, “well, this one has not turned down anybody, but this one is [a] little harsher of a grader.” I’m not sure how the system accounts for that so that we have consistency.22

In a similar vein, the subjective nature of some PTB tasks and the complexity of some tasks and exercises make it difficult to reach consensus when signing off:

Programs [are] complex. Most of our tasks are tied to the program and sometimes a little bit subjective. Even the program will change state to state based on the state’s needs, restrictions, statutory laws, cost share, and things like that.23

Particularly concerning are problems related to bias and accountability. We heard instances of PTBs being readily approved by friends and inexplicably rejected if a C&E found disfavor with trainees.24

21 GAO, 2020, p. 27.
22 Round 1 SME interview (#9).
23 Round 2 group interview (#10).
24 Round 1 group interview (#1).
Both GAO-20-360 and our interviews point to the absence of evaluative processes for C&Es directly contributing to problems such as bias and unaccountability.\(^{25}\) For example, one SME stated,

> [we] don’t evaluate the C&E at this moment, no. So I’m a C&E, I coach and evaluate you, but no one sits down with me and has a conversation about what I should be doing. There’s no follow up, [and] it’s something that could be done. I know it will require more people, but if we want to make it work, we need that.\(^{26}\)

**Training and Coaches and Evaluators**

Some SMEs feel that training for C&Es is problematic. One interviewee noted that the amount of required training for C&Es is quite limited: “we only have two days—16 contact hours—to convey information to a coach and evaluator.”\(^{27}\) Another interviewee stated that “a greater investment needs to be made in our coach/evaluators. Invest more time and more resources into developing effective coaching evaluation . . . we invest very little in trying to create effective people or people that can effectively conduct on-the-job training.”\(^{28}\) We also learned that some cadres are planning to develop more granular C&E training that provides additional guidance: “It is part of our goal to have a little bit more from our C&Es. We’re still developing that, and it is our intent to have more specialized [training] with our C&Es.”\(^{29}\)

**Assessing the Effectiveness of Coaches and Evaluators: Findings and Recommendations on What to Measure or Evaluate**

Interview responses showed that few if any official means exist to evaluate C&E performance or evaluate their adeptness at evaluating others. One interviewee noted that a C&E assessment process was briefly stood up before the coronavirus disease 2019 (COVID-19) pandemic but has not been implemented since.\(^{30}\) Other SMEs stated, “we haven’t really codified the agency’s overarching principles on coaching and evaluating . . . we haven’t really analyzed

\(^{25}\) GAO, 2020.

\(^{26}\) Round 1 group interview (#1).

\(^{27}\) Round 2 SME interview (#3).

\(^{28}\) Round 2 SME interview (#3).

\(^{29}\) Round 2 group interview (#10).

\(^{30}\) Round 2 SME interview (#19).
their effectiveness," and “what we’re doing now is trying to provide more evaluation of the C&Es endorsements and not just accept them because they’re C&Es.”

In Table 3.1, we recommend measures and related actions that FEMA can use to evaluate the effectiveness of C&Es.

Regarding some of the proposed metrics to address C&E rating quality, common errors in ratings include leniency and severity error (giving largely positive or negative ratings, respectively, in or across candidates), central tendency error (using largely the middle of the scale [e.g., a “3” on a 1–5 scale]), and halo error, in which an overall positive impression of a candidate influences leads to positive ratings across performance criteria that are not related to each other (this can occur with a negative impression and ratings as well). Regarding other investigations of bias, statistical analysis can show whether there are systematic differences in ratings as a function of T/Cs’ protected classes, such as gender, race, and age.

For performance review of C&Es to be effective, performance evaluation forms must accurately reflect C&E job requirements and have clear definitions of performance criteria and ratings options, and ratings must have sound psychometric properties (e.g., internal consistency, interrater agreement, and discriminability among levels of performance). Reviewers (typically supervisors) also must have opportunities to observe personnel doing their jobs. 360-degree evaluation can sometimes fill gaps because peers and subordinates might be in a better position to observe employees’ behaviors on the job. In general, most organizations

---

31 Round 2 group interview (#3).
32 Round 2 group interview (#9).
that use 360-degree reviews do so for developmental purposes rather than administrative decisions (e.g., pay raises or promotions), as reviewers might be motivated to give overly positive ratings (e.g., because they find it stressful to give negative feedback, worry that their ratings will not be anonymous, or feel pressured by the ratee to inflate the ratings) or overly negative ratings (e.g., to sabotage the ratee).33

The subjective nature of PTB tasks highlights the importance of having clearly defined criteria and rating options for C&Es to conduct their evaluations. In previous studies, we presented specific task statements along with such rating options as the following: trainee cannot perform the task, trainee can perform the task with assistance, trainee can perform the task with no assistance, and trainee can train someone else to perform the task.34 We discuss these measures in more detail in Chapter 5.

Course of Action Recommendations

The recommendations below were informed by the interviews and relate to CoAs that FEMA can adopt to improve the C&E program. They differ from those related to measurement and evaluation.

1. Through official FEMA memoranda, official emails from the administrator, IW-wide emails from senior leaders, and using training curricula: (1) publish and publicize updated and clear expectations for C&Es; and (2) publicize the critical role played by C&Es in maintaining the readiness and functionality of the IW, which might incentivize members of the IW to complete the requisite training and assume the role of C&E.

2. Create a mechanism for C&Es in the field to inform training personnel at FIWA about T/C readiness problems, which can be addressed by revisions to curricula.35 C&Es can be mandated to provide generalized feedback on T/C readiness—to inform approaches to training—via email before redeploying from the field.

3. Hire more C&Es and expand C&E training to focus on obscure PTB tasks and perform evaluations under challenging circumstances. One of several solutions might be increased role-playing or scenario-based training using real world examples of difficult PTB situations, which might boost the confidence of C&Es to sign off on a broad variety of tasks across cadres.36

33 See Hardison et al., 2015.
34 Lytell et al., 2017; Pulakos et al., 2002; Straus et al., 2014.
35 Round 2 SME interviewee #8 explained that in situations in which individuals repeatedly fail to independently demonstrate task proficiency, the C&Es should “come back to the training department and relay that information for use at future training Academy sessions.”
36 Note that this recommendation straddles both categories of C&Es and training.
4. If none currently exists, establish a grievance/redress process for trainees who feel unfairly evaluated by C&Es or neglected altogether and to address potential biases identified in analysis of C&E ratings, as described in Table 3.1.  

5. Deploy individuals with a dedicated C&E role and remove deployment-related duties outside their C&E responsibilities. Citing the success of the 2019 pilot program to streamline C&E duties (by solely focusing on coaching and evaluating), codify the new practice and publicize the intended outcomes (improved performance of C&Es).

---

37 A distinction should be made between feedback and grievances. A provider of feedback does not expect a specific response or resolution to occur. However, with grievances, the complainant seeks a direct response or redress. One of several ways to set up an effective redress mechanism is to designate an officer to receive, catalog, track, or monitor grievances. FEMA human resources officers can work with the Integration Branch or others to institute such a redress process.
CHAPTER 4

Training

Training is one of three pillars—in addition to experience and demonstrated performance—that contribute to the development of a skilled IW.1 FEMA's Field Operations Directorate is tasked with overseeing training and collaborating with various stakeholders—including cadres—to develop and deliver training to IW personnel.2 One stakeholder, FIWA, develops and provides training focused primarily to those in leadership positions.3 FIWA is also responsible for managing “courses that are cross cutting” across cadres, capstone courses, and certifying cadre instructors to teach the C&E course.4 Cadres develop and deliver the rest of the training to their personnel, with input from the FQS Branch on training curriculum development, resulting in the delivery of more IW training than that provided by FIWA.5 After training is delivered, the FQS Branch is responsible for conducting “course evaluation of all FQS courses by using the Kirkpatrick’s Four Levels of Evaluation (Level 1, 2, 3, and 4) as a standard.”6

In this chapter, we focus on training access and scheduling, collaboration and integration among different FEMA organizations, training evaluation, and learning management systems (LMS) and learning content management systems (LCMS). Chapter 2 discussed alignment of training and PTBs. Chapter 3 discussed training and C&Es—both in terms of training that C&Es provide and training that C&Es receive.

1 For this project, we focused on understanding position-specific training provided to IW personnel and to C&Es. FEMA’s Office of the Chief Component Human Capital Officer (OCCHCO) mandatory training for all FEMA employees is outside project scope. FEMA, 2022a.

2 FEMA, 2018b. Although this directive provides guidance over responsibility for training, we describe challenges we experienced in trying to understand training roles and responsibilities in the “Collaboration and Integration in Training Functions” section.

3 A Round 2 SME interviewee (#3) noted that the “Integration Branch is also responsible for the development of new leadership training that’s going to replace the stuff that FIWA currently teaches.”

4 Round 2 SME interview (#3).

5 “The percentage of incident workforce personnel training courses that are provided by FIWA versus the cadres is small,” according to a round 2 SME interviewee (#3).

6 FEMA, 2020, p. 17. Only Kirkpatrick’s Levels 1 and 2 are described in the FQS training procedures, with a note that a forthcoming FQS training evaluation standard operating procedure will provide more detail about evaluating training. (As of March 2, 2023, HSOAC had not received the standard operating procedure from the Integration Branch.)
Strengths of Training

As noted in Chapter 2, a strength related to training development appears to be FQS’s use of the curriculum mapping process in collaboration with cadres to ensure that training aligns with PTBs. Although there is not a way to evaluate alignment between training and PTBs as of the writing of this report (autumn 2023), some SMEs perceived alignment occurring between these two FQS sections.

One interview participant mentioned that changes that resulted from the COVID-19 pandemic led to improvements in how training is delivered. Before the pandemic, technical courses were taught in the field over five full days, whereas now, the courses are taught over eight days, with some half-day and some three-quarter-day sessions. The interviewee reported that this change gives the students “more time to let that information seep in,” which makes the students better prepared for the next lesson and enhances training efficiency.7

Areas for Improvement

Training Scheduling and Access

Several SMEs raised concerns about how training schedules and methods affect IW readiness. In our Round 1 interviews, some SMEs indicated the need for training before being deployed, noting the tension between FEMA’s need to deploy IW personnel and those IW personnel potentially being undertrained to perform key duties. One SME noted, “[We] need to see people go to a training course before they are deployed. It might also be valuable to send people out on training deployments to learn how to do the job, which is not something we do.”8 A different SME proposed a solution to helping IW personnel obtain experience in advance of deployments when they said, “I wonder if there’s some way to create realistic virtual reality (VR) environments where you don’t necessarily need to be deployed to a specific type of disaster but that type of disaster could be built into the VR [training].”9 This same SME noted,

[VR] keeps the training fresh, and it also appeals to a younger generation that would probably see that as something that they want to simulate. [Currently] I’ve got to go listen to somebody brief some old PowerPoints for six hours a day.10

Several SMEs also commented on the lack of refresher training and the lack of opportunities to deploy, which results in skill atrophy. One noted, “If someone goes through all the

---

7 Round 2 SME interview (#6).
8 Round 1 group interview (#11).
9 Round 1 group interview (#9).
10 Round 1 group interview (#9).
training, they might be able to do the task when they get signed off, but three or four years go by, and they don’t have to do a deployment, so they forget how to do it.”

Interview participants also commented on access to training. One SME stated that meeting the demand is not a problem for the courses that their cadre owns, but it can be challenging to get personnel in courses owned by other branches or cadres, often because those entities do not have enough instructors to staff the courses, courses are not offered frequently, or a cadre prioritizes access to its own members over those from other cadres: “Most of the time I have to beg and plead to get spots.” Other participants noted that funding—and even such factors as availability of classroom space—limit the number of instructors and students who can participate in each course.

A participant responsible for training delivery discussed challenges in determining and meeting demand, because different sections in FIWA are responsible for different aspects of training management and delivery:

> What is the actual training demand? I asked for more instructors even though I don’t really have a goal on throughput, I just know that it’s not enough. . . . It’s very challenging to determine whether or not our scheduling meets the requirement. We’re trying to make sure we’re asking the right questions and working with the cadre to make those determinations.

In both Round 1 and Round 2 interviews, several SMEs remarked that not being able to enroll in training courses prevents personnel from completing PTBs and generates a lot of frustration; for example, one participant stated, “I bet there’s one class I have needed for two years, it hasn’t been offered or got cancelled or whatever or I couldn’t go that particular week, and that’s what’s stopping me from completing my PTB.” Another SME noted that FEMA has not delivered an Operations Chief course since 2009: “I’m guessing there are items on the PTB that try to assess proficiency in that regard, but the courses aren’t there to help train individuals on it.” Compounding the issue of access to training is that IW personnel tend to prioritize getting deployed in their position as opposed to attending training for that position, which results in more than “50 to 60 percent of potential trainees [skipping training] because they’re more focused on getting to a disaster instead of actually coming and getting trained on their specific job.” When this happens, the T/C might be unprepared for the disaster, particularly if they are new to FEMA and their position because, as a SME noted,
“they’re getting deployed to a disaster when they don’t have the basic skills to actually know what they’re doing.”

Collaboration and Integration in Training Functions

Several SMEs commented on integration among entities responsible for training development, delivery, and evaluation, with varying and conflicting views on the optimal degree of collaboration. In addition, we found it difficult to pinpoint which branches or sections are responsible for different aspects of training based on a review of GFI in combination with SMEs’ comments about their roles in training development, delivery, and evaluation. For example, although the FQS Training Procedures (FEMA, 2020) state that the FQS Branch is responsible for conducting Kirkpatrick Level 1 through 4 evaluations, SMEs in our interviews reported that the cadres were responsible for conducting evaluations. One SME highlighted these issues in the following remark about course development:

> Training is divvied up across the entire agency. . . . FQS is its own entity but doesn’t have rights to that. It’s all divided across the cadres and the cadres don’t even all fall under the same types of leadership. Some of them fall under one part of the agency [and] some of them fall under other parts of the agency. There’s no one singular entity that owns the guidance for all development.18

Some SMEs advocated for more integrative and collaborative efforts. For example, a SME from an HQ branch remarked that it is unusual that although FIWA delivers most of their curriculum, they are not necessarily involved in creating it. Another SME had a similar comment regarding the value of a collaborative effort, because personnel who deliver training are aware of factors that affect feasibility of implementing it:

> From a delivery perspective, [we can] ask those questions to the overall body that’s developed the curriculum, and say, “hey, how exactly you are going to do that in real time and try to make sure that things stay on course?”19

A SME from a cadre discussed the value of having more multi-cadre course development or information-sharing to make use of other cadres’ materials when there is overlap among content: “for example, when external affairs is talking about what operations does, they should not be creating those slides from scratch. They should be coming to the operations cadre.”20 According to another SME from a HQ branch, “It just seems to me that there would be some value in standardization or collaboration across the cadres in the kind of

17 Round 1 group interview (#4).
18 Round 2 SME interview (#3).
19 Round 2 SME interview (#3).
20 Round 2 group interview (#12).
Training they need or what they need training on with respect to FQS. There would be overlap or consistency and needs.” 21 However, this SME also noted, “That's definitely stuff that's out of my lane.” 22

A different noncadre SME had opposing views with respect to collaborating on course delivery and assessment efforts: “We're in the delivery business. My instructors became course managers just out of necessity. It’s not really their area of expertise and frankly they don't have the bandwidth to delve into the broader responsibilities of course management.” 23

This SME pointed to the tension between collaboration and silos or stovepiped efforts,

We have so much work to do that it's almost impossible to not have stovepipes. We try so hard to not have stovepipes, but we still have to in order to just kind of get things done and keep things moving forward. You can end up collapsing under so much collaboration because you can't get decisions made.24

Learning Management System and Learning Content Management System

Several interview participants also addressed the need for an LMS or LCMS to support training. Currently, DTS is the system of record, but DTS was designed as a deployment tool. With respect to training, DTS is limited largely to documenting whether an FQS student receives credit for a course, and it can store attachments for administrative purposes (e.g., a spreadsheet with students’ test scores), but it does not have the functionality of an LMS. FIWA uses two Emergency Management Institute products to store FQS training content, but this is for storage purposes only.

Several interview participants discussed the fragmented nature of information technology (IT) systems to support training; for example, participants reported the use of multiple systems—including DTS, SharePoint for providing access to content (but not for content creation), use of the Emergency Management Institute’s LMS to store FQS content, Scantron sheets for test administration (which shifted to Microsoft Forms for virtual instruction during COVID-19), Survey Monkey or Microsoft Forms for Level 1 surveys, and instructor portals that are being discontinued because they are no longer IT compliant. However, these systems do not talk to each other. As one interview participant stated, “We have various systems. Each one does a tiny little part to support FEMA qualification, but we don’t have one that does the important work for curriculum.” 25

21 Round 2 SME interview (#20).
22 Round 2 SME interview (#20).
23 Round 2 SME interview (#3).
24 Round 2 SME interview (#3).
25 Round 2 SME interview (#3).
In addition, there are six different LMSs in use across FEMA entities, but they are not integrated with each other or with DTS for reporting course credit. FEMA has a goal to adopt an enterprisewide integrated LMS (FEMA Integrated Learning System [FILS]) with a wide variety of functions to support training, as described in a 2021 Capability Analysis Report.\textsuperscript{26} Interview participants’ comments about the need for FILS capabilities point to weaknesses or gaps in FEMA’s current training development, delivery, evaluation, reporting, and management functions. Interview participants indicated that FILS should have the following capabilities:

- content creation and revision, including automatic updating (e.g., “You can make that change one time instead of having to review thousands of pages, just to change someone’s name.”)\textsuperscript{27}
- content delivery
- access by multiple stakeholders (content creators, instructors, students, administrators), along with access controls
- communication
- sufficient bandwidth for storage and streaming, given that a lot of material used video
t- training evaluation (including a function that automatically sends Level 1 surveys to students for completion, with results reported to cadre leadership via a dashboard)
- registration.

Some interview participants emphasized the need for a system similar to what colleges and universities use: “One of those systems would push us light years ahead.”\textsuperscript{28} However, FILS has not been funded as of April 2023.

At the same time, interview participants identified several challenges in developing or adopting a single system. In addition to funding, a concern is that FEMA might need customized functionality that might not be available in off-the-shelf products; as an example, one interview participant noted, “If someone fails a task in the field, it can be linked back not just to a specific training course but to a specific module day and time. I think that would be

\textsuperscript{26} FEMA’s Capability Analysis Report for FEMA Integrated Learning System (FILS) Requirements and Acquisition Analysis (2021) identified 83 necessary capabilities for an LMS/LCMS across seven tasks: system administration, content management, content delivery, assessment, communication, analysis and reporting, and cross-cutting and supporting attributes. Seventy percent of these 83 capabilities have "some form of system gap" (p. vi). The Capability Analysis Report also identified several possible solutions for an enterprisewide LMS/LCMS; these include acquiring a commercial-off-the-shelf or government off-the-shelf product, developing application program interfaces to enable select data sharing across existing LMSs or LCMSs in use, expanding one of the existing systems to other FEMA organizations, or acquiring a custom-built system (FEMA, 2021).

\textsuperscript{27} Round 2 SME interview (#21).

\textsuperscript{28} Round 2 SME interview (#20).
important for that LMS to be able to do.”

Perhaps the most difficult challenges are identifying a system that meets all entities’ needs and the willingness of the entities with existing LMSs to switch to a different system. As one participant said, “No one is willing to let theirs go.”

Assessing the Effectiveness of Training: Findings and Recommendations on What to Measure or Evaluate

In contrast to other FQS elements, SMEs talked extensively about metrics to evaluate training effectiveness. Responses focused on Kirkpatrick Level 1 through 4 evaluations and the need for an LCMS to administer and record data for training evaluation, as described previously.

Kirkpatrick Levels 1 Through 4

FEMA uses Kirkpatrick’s four-level framework, depicted in Figure 4.1. In brief, FEMA conducts primarily Level 1 evaluations. Some branches or cadres also administer tests to assess Level 2, but such efforts appear to be sporadic.

FIGURE 4.1
Kirkpatrick Levels of Training Evaluation

1. Reaction
   How satisfied are participants with training?

2. Learning
   To what extent do participants’ knowledge and skills improve as a result of training?

3. Behavior
   To what extent do participants apply what they learned during training when performing their jobs?

4. Results
   To what extent do organizational outcomes improve as a result of training?

SOURCE: Adapted from Kirkpatrick and Kirkpatrick, 2016.
Level 1

Responses from SMEs indicated that FEMA conducts primarily Level 1 evaluations, but those efforts need improvement. For example, a SME stated that Level 1 end-of-course evaluations need to be redesigned to be “more methodical and [to] provide feedback in a more comprehensive way so we can look at it by cadre, by course, by instructor, etc., so we can get better feedback to the courses.” For example, although all students are supposed to get the same set of questions (which are stored in a Word document), if cadres decide to change the questions, HQ entities would not know. Some cadres also reported collecting open-ended comments or feedback from students via interviews or after-action reviews, although these efforts are sporadic.

Another gap in Level 1 assessments pertains to how data are collected and used. Before COVID-19, Level 1 surveys were administered in person, using Scantron sheets, but post–COVID-19, a variety of methods were used for survey administration, and no one was responsible for collecting and analyzing the data: “It’s all over the place right now.” This SME added, “Cadres are trying to do that on their own, but what are they doing with it? It’s not collected in a central location.”

Another interview participant reported that these changes in administration of Level 1 surveys were not a result of the shift to virtual instruction but occurred when WDD ended its partnership with the Emergency Management Institute, which had previously managed this activity. This SME felt that that WDD still needs to “get their sea legs under them on these processes and formalize them.” Furthermore, the SME reported that turnover in WDD has hindered development of Level 1 evaluations because when personnel depart, the division loses background information about the plans: “we’re constantly just restarting . . . because there is a change in personnel.” This suggests that the division might lack a mechanism to maintain institutional memory. Furthermore, this SME stated that interruptions in planning are exacerbated when a big event occurs: “It’s an all hands-on-deck to the response, and then you’re coming back and starting over six months later.”

---

31 Round 2 SME interview (#20).
32 A SME from one cadre described conducting interviews following training exercises, asking such questions as “What did you learn in this class? How have you used that [task] in the field? How have you used [it] with your position?” A SME from another cadre described evaluations that they conduct in the first year of new courses whereby the top section chief and/or branch director sits in on part of the class and conducts an after-action review with students at the end of the course.
33 Round 2 SME interview (#20).
34 Round 2 SME interview (#20).
35 Round 2 group interview (#10).
36 Round 2 group interview (#10).
37 For example, see Carley, 1992.
38 Round 2 group interview (#8).
The lack of centralized collection or analysis of Level 1 data inhibits use of findings to make improvements across cadres or other entities, and there are challenges to applying results even in cadres. As one SME reported, “Every once in a while, we will make some tweaks and changes.” Another SME stated, “We haven’t been in a place where we could even truly implement some of the feedback that we received, though we tried to.” Timing of feedback is also problematic, as noted by a SME who stated, “If you ever saw any feedback, it was always months later either as an instructor or a course manager—and the feedback that was shared was “lacking” from a content perspective and not “substantive.”

Level 2

Stakeholders recognize the need to conduct Level 2 assessments but identified barriers to doing so. For example, the first SME cited in the preceding paragraph stated that improvements to training “should be steered by legitimate Level 2 feedback, but we don’t have those mechanics.” A SME from a HQ branch reported that they are not resourced to go beyond Level 1 evaluations to evaluate training for C&Es. A SME from another HQ branch noted that FEMA has a goal to focus more on Level 2, but it is currently up to the cadres to assess learning. The division would also like cadres to implement more hands-on assessments of behaviors in capstone exercises rather than written tests (and they are in the design phases of trying to make that happen) or have a C&E evaluate personnel in the field to determine whether they can perform tasks.

There have been some attempts to conduct Level 2 evaluations. SMEs from three cadres we interviewed reporting using pre-training and post-training knowledge tests or post-training knowledge checks. For example, one cadre SME stated that they are required to conduct a pre-test and post-test to assess improvements in learning. The tests can consist of the same questions, or the post-test can assess the same content with reworded questions in a randomized order. Another cadre also reported using pre-course and post-course exams to gauge what the participants have learned. The SME stated, “that’s kind of the best tool that we have [to assess training effectiveness].”

Furthermore, in our discussions of using tests for Level 2 evaluation, it appears that student performance on tests and in the course overall is recorded as pass/fail. In comparison to using continuous scores (e.g., zero to 100), pass/fail grades restrict variability in performance

---

39 Round 2 SME interview (#3).
40 Round 2 group interview (#10).
41 Round 1 group interview (#3).
42 Round 1 group interview (#3).
43 Round 1 group interview (#3).
44 Round 2 SME interview (#3).
45 Round 2 group interview (#10).
data and limit what organizations can learn about training effectiveness. In addition, one SME described a lack of integration and expertise in terms of developing tests that map well to curricula and analyzing test quality:

If I'm writing a test, then, ideally, I've got some data as a curriculum that's helping me determine how that test or activity is structured in order to determine whether learning is taking place. I guess in the absence of that we're going to do the best we can and we're going to create a test with an expectation of how each test item is going to perform. When that test is fielded then we do an analysis of the data that comes back and say, "hey, you know what, test item number 2 [is] all over the place and [there] must be something wrong with the way we worded this test question." . . . Maybe it doesn't match up with the supporting materials as delivered or maybe the instructor notes are vague. So, the instructors don't understand how or what they're supposed to be teaching.

This SME added,

We don't really assess the performance of the test itself. I mean, we just take actions based off of whether [the students] graduated or whether they need remediation . . . . That's the huge problem with not [having] a Level 2 mechanism. I mean, we have tests, but the analysis isn't happening, and that's a resource issue in terms of having people to analyze that data.

Levels 3 and 4
Several SMEs stated that FEMA is not conducting Level 3 and Level 4 evaluations. For example, one of the cadre personnel who discussed use of pre-tests and post-tests stated, "We don't have any real mechanism or [post-test] from the field to know [whether] what they learned was on target." However, one SME stated that the Incident Management Exercise (IMEX) program, a relatively new simulation, can “bridge the gap between training so that IW get an actual deployment experience.” Similarly, a SME pointed out that IMEX might provide an opportunity to collect Level 3 data.

---

46 Even when using continuous scores, it is important to have scores that discriminate among levels of knowledge and skills. If most students obtain high (or low) scores, the test will not distinguish among levels of knowledge and skills.

47 Round 2 SME interview (#3).

48 Round 2 SME interview (#3).

49 Round 2 group interview (#10).

50 Round 2 SME interview (#7). We recognize that these two SMEs characterized exercises as training, a perspective that might differ from other IW personnel or Integration Branch management.
Another SME remarked, “I don’t think FEMA is ready to build a Level 3 because we don’t have a good handle on Levels 1 and 2 yet.” This SME noted that Level 3 evaluations are essential to curriculum builders and personnel in the field to get input on whether they are prepared to do “what they need to do.” This expert also commented on the importance of developing courses and writing tasks in ways that lend themselves to Levels 3 and 4 evaluations (e.g., “[Tasks] hadn’t been written objectively or with an outcome in mind, so we have contractors looking at that”). Given that data for Level 3 evaluations typically serve as inputs for Level 4 analyses, it is not surprising that FEMA has not attempted Level 4 evaluations.

Assessing Training Scheduling and Access
Comments regarding measures to assess the match between training demand and supply were mixed. For example, a SME stated that FIWA conducts a needs analysis to determine how many people need to take a course (presumably using DTS) to decide how often to offer courses, creates a draft training calendar, and sends it to cadres for review. However, the SME also reported that cadres are often too busy to review the calendars in detail, and therefore they later request more courses or report that they do not need courses that are being offered at particular times. As discussed previously, a SME referred to the complexities of determining whether scheduling meets requirements because of the involvement of multiple sections of FIWA in training management and delivery.

Table 4.1 recommends measures and actions to improve Levels 1 and 2 evaluations and to assess training scheduling and access.

A standardized set of questions for Level 1 surveys would allow for comparisons across courses and over time. Customized questions can allow cadres to add items reflecting their unique content or instructional methods and obtain qualitative feedback in response to open-ended questions. In addition, we recommend sending automatic “triggers” for trainees to complete the Level 1 evaluations to receive course credit (as one SME suggested) and use of automating scoring of Level 1 closed-ended questions and dissemination of results to stakeholders (e.g., via a dashboard) in a timely way.

For surveys assessing student self-efficacy, Table 4.2 shows an example of a set of items used in a RAND study of effectiveness of intelligence analyst training. The column labeled “Proficiency” was used to assess self-efficacy. Trainees answered the questions for each of approximately 24 tasks. Trainees completed the assessment at the beginning and end of train-

---

51 Round 2 SME interview (#20).
52 Round 2 SME interview (#20).
53 Round 2 SME interview (#20).
54 Round 1 SME interview (#5).
55 Round 2 SME interview (#3).
Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions

Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions

We note that this measure reflects attitudes about or affective reactions to the course—it is not a measure of actual learning, which would be assessed in Level 2 evaluations.

Ideally, Level 2 evaluations of learning would consist of tests of both cognitive knowledge and behaviors or skills. Pre-test and post-test designs provide opportunities to assess improvement in knowledge and skills, but it might not be practical to administer some behavioral assessments prior to training (e.g., trainees might not yet know how to operate equipment that they will be trained to use). Behavioral assessments can also be more costly to

### TABLE 4.1
**Recommended Measures and Actions to Assess the Health of Training**

<table>
<thead>
<tr>
<th>Gap</th>
<th>Recommendations on What to Measure or Evaluate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability of classes</td>
<td>• Continue needs analysis and identify strategies for increased cadre involvement</td>
</tr>
<tr>
<td>Standard and systematic Level 1 evaluations</td>
<td>• Create a combination of standardized questions and questions that cadres can customize that are administered online</td>
</tr>
<tr>
<td></td>
<td>• Administer online assessments of pre-training and post-training self-efficacy for task performance&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Quality of Level 2 assessments</td>
<td>• Conduct practical exercises (work samples) that reflect the job requirement to assess behavioral outcomes during training, using well-defined checklists and clear options for rating proficiency</td>
</tr>
<tr>
<td></td>
<td>• Create carefully constructed written tests of substantive knowledge for course concepts</td>
</tr>
<tr>
<td></td>
<td>• Create pre-tests and post-tests to assess improvement in learning</td>
</tr>
<tr>
<td></td>
<td>• Analyze the psychometric properties of tests, ideally using item response theory&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>• Retain the full range of test scores, not only “pass/fail” outcomes, to assess training effectiveness</td>
</tr>
</tbody>
</table>

<sup>a</sup> See Lytell et al., 2017, and Straus et al., 2014, for examples of question formats for self-efficacy. Each cadre would need to populate the task statements in this format.

<sup>b</sup> Item response theory (IRT) is a mathematical modeling approach used to construct, assess, and score instruments that measure individual knowledge, skills, abilities, and traits. Unlike classical test theory, which provides information about the psychometric characteristics of the instrument as a whole or treats each item as equivalent, IRT provides information about each item’s characteristics (e.g., difficulty or probability of a “correct” response) according to the relationship between individuals’ responses to the item and their overall ability or trait that the instrument intends to measure.

### TABLE 4.2
**Sample Item from Lytell et al., 2017**

<table>
<thead>
<tr>
<th>Task</th>
<th>Familiarity</th>
<th>Frequency</th>
<th>Proficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Define the operational environment</td>
<td>☐ familiar</td>
<td>☐ never</td>
<td>☐ I cannot perform this task</td>
</tr>
<tr>
<td></td>
<td>☐ not familiar</td>
<td>☐ a few times a year</td>
<td>☐ I can perform this task with assistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>☐ monthly</td>
<td>☐ I can perform this task with no assistance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>☐ weekly</td>
<td>☐ I can perform this task with no assistance and I can train someone to perform this task</td>
</tr>
<tr>
<td></td>
<td></td>
<td>☐ daily</td>
<td></td>
</tr>
</tbody>
</table>

**SOURCE:** Reproduced from Lytell et al., 2017.
conduct if they require instructional staff to serve as raters. Written tests that affect administrative decisions, such as pay or promotions, might also create potential legal issues if not properly validated, as noted by a SME in a Round 2 interview.56

We agree with SMEs that FEMA might not be ready for Levels 3 and 4 assessments, but we have some general recommendations about initial steps. First, we recommend that FEMA continues efforts to write objective task statements and develop curricula to prepare for Level 3 and 4 evaluations, as one SME described. These efforts can also contribute to developing Level 2 tests, and, according to some SMEs, contractors are developing curricula and knowledge tests in concert. Second is to consider feasible designs for Level 3 and Level 4 studies. Although there are multiple approaches to conducting these assessments, two general approaches to evaluating transfer of training to the job (Level 3) and organizational results (Level 4) that provide the strongest evidence are predictive validity studies and controlled experiments. A predictive validity study involves calculating the correlation between performance in training with performance on the job (Level 3) or other relevant disaster-related outcomes (e.g., cost, efficiency, effectiveness; Level 4). A positive correlation indicates that individuals who perform better in training also perform better on the job, and individuals who perform worse in training perform worse on the job (or that higher or lower performance scores in training are associated with better or worse organizational outcomes, respectively). This approach requires having psychometrically sound measures of both performance in training and job performance (for Level 3) or organizational results (for Level 4). A controlled experiment entails analyzing differences in outcomes (performance or organizational results) between an intervention group that receives the training and a control group that does not (ideally, with random assignment of personnel to conditions). Understandably, some organizations resist using experiments because they do not want to withhold training from personnel.57

56 Round 2 SME interview (#20).

57 If there are personnel performing their jobs in the field who vary in completion of particular courses, and this variation is not associated with other systematic differences among these personnel, this might offer a natural quasi-experiment to evaluate training impacts of those courses. A pre-test and post-test can be combined with an experimental approach to assess whether improvement in knowledge and skills differs for the intervention and control groups. However, there are other challenges in conducting experiments, including the need for a large sample of personnel in each, availability and willingness of supervisors to rate T/Cs’ performance in the field if archival performance measures are not available, and potential awareness of personnel about the experimental conditions, which can influence results in unintended ways.
Course of Action Recommendations

The recommendations below were informed by the interviews and relate to CoAs that FEMA can adopt to improve training. They differ from those related to measurement and evaluation.

1. Provide training, training aids, and access to test-development experts to assist test developers if analyses of the quality of knowledge tests indicate the need for improvement (for example, the tests do not discriminate among levels of knowledge or behavior).

2. Develop and implement policy that establishes clear responsibilities and ownership of the various aspects of training—including development, delivery, and evaluation—so that there are clear roles across various FEMA stakeholders. Delineating roles and responsibilities can lead to better collaboration and coordination across stakeholders.\(^{58}\)

3. Investigate options, such as individual, simulation-based training, to maintain skill proficiency during long periods of time between deployments for some personnel.

4. Adopt an enterprisewide LCMS that maximizes needs, as identified in the 2021 Capability Analysis Report.\(^{59}\) However, given that this is not likely to occur in the near future, we recommend exploring the feasibility of purchasing added features of DTS that allow for data capture, storage, analysis, and dissemination.

5. Seek fiscal resources for sufficient FEMA and/or contractor personnel—and personnel with appropriate expertise—to increase course availability and support development, analysis, and use of metrics.

\(^{58}\) Gratton and Erickson, 2007.

\(^{59}\) FEMA, 2021.
As illustrated in Figure 5.1, the Qualification Review Board (QRB) plays a role—although it is not the ultimate authority—in whether a T/C with a completed PTB is determined to be qualified for a position. SMEs make up the QRB and receive several data inputs about the T/C to consider when reviewing for qualification: completed PTB, cadre review, and Certifying Official review. In addition, the QRB also poses questions to the T/C about the position for which they are seeking qualification. Tier 1 candidates are interviewed via in-person meetings, which are scheduled at least quarterly, to discuss several topics, including their deployment and wing history. Tier 2 candidates are reviewed via virtual meetings that

FIGURE 5.1
Role Played by the Qualification Review Board in the Trainee/Candidate Qualification Process

are held monthly. QRBs decide on a recommendation about qualification and send the recommendation to the CA, who ultimately decides about T/C qualification (i.e., the CA can agree or disagree with the QRB’s recommendation). SMEs explained that there are currently 38 QRB members in total—34 are from cadres, four are from Field Operations Directorate leadership—and that 19 of the 23 cadres are represented, with “four cadres not having QRB representatives because they do not have Tier 1 and Tier 2 personnel.”

**Strengths of Qualification Review Boards**

SMEs recognize that the intent of the QRB is to mitigate bias in the determination of a T/C’s proficiency. One SME noted, “I like the QRB concept because it’s like a peer review—especially for the Tier 1 leaders.” Other SMEs described the QRB as “an extra layer/filter to evaluate candidates outside the cadres [so that there is a] recommendation beyond the cadre” and “an independent body that is kept at arm’s length . . . so that they could provide an independent evaluation of each candidate.”

Some SMEs also discussed how the QRB serves as a checkpoint to ensure that a candidate meets all of the required benchmarks before their packet is sent to the CA. One described the QRB as a method to identify where a candidate’s packet can be strengthened by providing recommendations (e.g., further training or additional deployments). This SME reported that the QRB provides a “full list of findings and a report of the candidate [to the cadre].” Another SME noted that the QRB process prompts a thorough review and preparation of the candidate by the cadre: “I’ve never had [a candidate not be recommended by the QRB]. If I’m sending somebody to the QRB, we’re doing QRB prep so that they’re actually ready to do it.”

Finally, some SMEs described recent changes to the QRB that strengthen its utility. One mentioned, “what they’re doing now with the interview piece is really good.” Another SME reported that there is a lot of effort being put into updating the QRB process for it to be useful for the cadres. This SME believes that an increase in QRB members over the past few years indicates that cadres feel more ownership of the process:

QRB is updating everything, updating all the documents and providing services to the cadre that the cadres didn’t have before. I think cadres are more comfortable [with the

---

2 Round 2 SME interview (#18).
3 Round 2 SME interview (#10).
4 Round 2 SME interview (#18).
5 Round 2 SME interview (#11).
6 Round 2 SME interview (#18).
7 Round 2 SME interview (#6).
8 Round 2 SME interview (#10).
Areas for Improvement

Many SMEs expressed the sentiment that the QRB adds little value—or in some cases can be a hindrance—to the qualification process. In some interviews the topic elicited fairly strong reactions, with some SMEs using phrases such as “a waste of time . . . I don’t see a real value in the QRB,”10 “I think it’s a waste . . . I can make that decision by myself,”11 “has no value add,”12 and “I think it is ineffective and useless . . . an unnecessary extra step.”13 One common perspective is that the QRB does not add value to the process because, regardless of the QRB’s recommendation, the cadre CA holds final decision authority on whether a candidate is qualified.14 As one SME explained, “even if the QRB does not recommend someone for qualification, I will more than likely always go to my CA and have that person administratively qualified.”15 Another stated that if their cadre expects the QRB to push back on a candidate, “we don’t even bring it to the QRB anymore and we’ll bypass them by going to our CA for qualification.”16

One SME mentioned that they feel that the QRB adds little value because it does not provide sufficient feedback to improve a candidate’s qualification packet: “the reason that we have not found the QRB useful recently has been they will recommend nonqualification and provide no actionable information on what the candidate can do to improve.”17 This statement seems to be in direct conflict with the SME we previously mentioned who reported QRB providing a “full list of findings and a report” to the cadres.18 It is unclear why this discrepancy exists and where it originates.

9 Round 2 SME interview (#18). However, the view that the QRB process is improving was not shared by everyone. One SME commented that they have not felt an improvement in the QRB process despite recent personnel changes: “there’s a new person that runs the program now, but there was no standard change. It’s just a new group of people on the QRB” (Round 2 SME interview [#11]).
10 Round 2 SME interview (#6).
11 Round 2 SME interview (#7).
12 Round 2 group interview (#11).
13 Round 1 SME interview (#5).
14 Round 2 interviews (#5, #6, #10, #11).
15 Round 1 SME interview (#5).
16 Round 2 group interview (#11).
17 Round 2 SME interview (#11).
18 Round 2 SME interview (#18).
SMEs also reported cases where they felt that the QRB was actually a hindrance to the qualification process. Many of these centered around the question of whether QRB members have the correct expertise to judge the proficiency of cadre candidates. One SME stated, “we have a lot of people who are on the QRB who do not understand what [our cadre] does.” There are also some challenges centering around inconsistencies in official FEMA guidance. For example, there is no requirement in the FQS Guide—which is followed by the cadres—for some positions to require a deployment in order for a candidate to be qualified in that position; however, the QRB Handbook—which is followed by the QRB—requires an explanation to be provided if the candidate has not deployed in that position. Three separate individuals expressed frustration with this inconsistency.

One SME felt that the QRB process is a hindrance to FEMA’s goal of increasing the number of qualified responders in the field because it adds too much time to the qualification process:

We end up spending a lot of time, we wait a long time, for those people to get through QRB because they only have it once a month (sometimes only once a quarter) and to get through those cycles can take a considerable amount of time, and we have to do waivers to explain why our cadre doesn’t deploy in our candidate titles, because it’s considered a liability to put someone who is not “qualified” out in the field.

Finally, in one instance, a SME indicated that in practice, QRB recommendations are not always impartial and unbiased. This individual described an incident where differences in perspective about job procedures between a candidate and a QRB member—rather than a legitimate shortcoming in the candidate’s qualification packet—resulted in a QRB recommendation that the candidate not be qualified for their position.

Assessing the Effectiveness of the Qualification Review Board: Findings and Recommendations on What to Measure or Evaluate

According to discussions with SMEs, the largest challenge facing the QRB process is the lack of trust in the value and legitimacy of QRB recommendations. There appear to be no existing metrics for FEMA to gauge QRB effectiveness; however, several measures could be used to address this issue.
First, FEMA could assess perceptions of QRB stakeholders by conducting regular, brief online surveys. These surveys could include questions to gauge perceptions on, for example, the overall value of the QRB process; whether cadres prefer to bring candidates directly to the CA instead of engaging the QRB; whether the QRB provides relevant, actionable feedback; whether the QRB has sufficient expertise to judge a cadre candidate; whether the QRB process is timely; and whether the QRB is equitable and impartial.

To measure equity in QRB decisions, FEMA could capture data (if they do not already exist) on QRB recommendations by cadre, protected class, or other factors. Examining patterns that might emerge over these demographics can help illuminate conscious or unconscious biases in the QRB process and outcomes. Addressing apparent inequities and sharing results from these analyses with QRB stakeholders might increase the perceived legitimacy of QRB recommendations.

Finally, adopting a quantitative model could make QRB qualification recommendations more consistent, further increasing their perceived legitimacy among stakeholders. There is robust literature and substantial evidence that equations (also referred to as mechanical methods, actuarial judgments, or linear models) produce equal or better outcomes compared with human judgment (also called subjective judgments or clinical judgment) in combining inputs to make decisions, even when the human judges provide ratings or scores of the inputs used in the equation.24 The consistency that such a model provides enhances the validity of decisions. The use of such a model—along with transparency about its use—might enhance stakeholders’ perceptions of the legitimacy of QRB decisions. One difficulty in adopting a standardized quantitative model will be ensuring that it sufficiently captures nuance in the differing needs of each cadre.

To use a quantitative model, QRB members would independently assign a numeric indicator (with well-defined response options) to each input (unless inputs already have numeric ratings). Scores would be aggregated across QRB members and combined into an equation (e.g., summed or averaged), with a predetermined cutoff score that determines whether a T/C is qualified.25 Different weighting factors can also be considered: Each input could be equally weighted, or input weights could be determined in advance by SMEs in a separate exercise, or—the best-case scenario—weights could be determined statistically by analyzing the association of input scores with measures of T/Cs’ job performance (this option requires that FEMA has good measures of T/C performance).

Table 5.1 summarizes these recommendations.

---

24 Dawes, 1979, p. 571; Grove and Meehl, 1996, p. 293.

25 Aggregate scores on each separate input could also be analyzed to determine whether members agree on rubrics used to score the inputs.
Course of Action Recommendations

The recommendations below were informed by the interviews and relate to CoAs that FEMA can adopt to improve QRB processes. They differ from those related to measurement and evaluation.

1. Conduct analyses to determine the return on investment (ROI) of the QRB. An ROI study should balance considerations of perceptions of utility, cost of QRB member and candidate time, impartiality of recommendations, timeliness, and potential conflicts with other FEMA goals.
2. Consider an alternate QRB model that better addresses the technical nuances of each cadre and job position. For example, this could take the form of convening, for each cadre, a technical review panel with specialists specific to that cadre.26
3. Continue implementing changes in the QRB process that focus on providing additional utility to the cadres. Current QRB leadership is updating documentation and providing additional services to cadres to encourage participation and deepen cadre investment in the process; however, not all cadres have noticed an increase in utility. Continuing these efforts might extend added value to additional cadres and increase the perceived and actual ROI of the QRB.
4. Update the QRB Handbook to reflect that the FQS Guide does not necessarily require a candidate to deploy in a position for which they are seeking to be qualified. Allow the QRB to recommend nonqualification on the basis of no deployments (or require a nondeployment waiver) only for those positions that actually require a deployment.

---

26 This recommendation is based on a comment from one of the Round 2 SME interviews (#11) that compared FEMA’s QRB with the QRB of the National Incident Management System (NIMS): “If you read the documentation for what a QRB is under NIMS, it’s a group of technical specialists at a position that can actually perform [the tasks that are being evaluated], and [can therefore] evaluate a candidate’s ability to perform the tasks for a particular position. It’s not whatever we have at FEMA. Essentially [each cadre would have] our own QRB or technical review panel.”
CHAPTER 6

Deployment Tracking System

DTS is an electronic system that allows FEMA to “track and manage FQS, deployment and event records, and workforce readiness.”¹ DTS consists of two parts: (1) Deployer, which allows leadership to manage event planning and employees, and (2) Responder, which allows employees and supervisors to accept deployment opportunities and monitor PTB completion progress.² IW personnel who are deemed qualified for a position related to the QRB process described in Chapter 5 receive notification from DTS that the PTB under consideration was approved, and the approved PTB is removed from the Responder Portal.

Strength of the Deployment Tracking System

The key strength related to DTS concerns the fact that it contains a variety of potentially useful information.³ A SME familiar with querying DTS commented about the ability of the system to contain and extract valuable information: “DTS can give us answers to many questions, but we need to make sure the questions we ask are the right ones. In certain cases [if people are not getting the answers they want] it’s about what people are asking.”⁴ A different

² FEMA, 2018a. Going forward, DTS might benefit from a comparative analysis of similar deployment tracking or accountability processes to learn about best practices, problematic areas, and more. An expansive deep dive on this is beyond the scope of this study, but a cursory look suggests that parallels might exist between DTS and processes such as the U.S. Army’s Deployment Theater Accountability System (DTAS). According to its website, DTAS “provides reliable, timely, and efficient accountability for Soldiers, Marines, Airman [sic], Sailors, DoD Civilians, contractors and foreign nationals.” It also gives commanders “at all echelons the ability to track their personnel by name, unit, location, and date.” See U.S. Army Human Resources Command, 2023. Also, the U.S. Air Force’s Personal Deployment Processing Tool allows Airmen to review and track their readiness status for medical and personnel requirements.
³ Although we reviewed material that described the Deployer and Responder portals in DTS, SMEs did not discuss these portals specifically as strengths or areas for improvement.
⁴ Round 1 SME interview (#6).
SME noted that DTS seems effective at what it was designed to accomplish, though the original design might be changed if the system was redesigned today:

It’s like building a little house, and then you have 12 kids, and you put a new room on every time you have a kid. It’s good enough at housing the 12 kids, but I think that if we all sat down today and designed a new system, we would have a very different vision of what our ideal state was.5

Areas for Improvement

We learned from Round 2 interviews that DTS contains a lot of data, but the quality and meaningfulness of the data are uncertain. SMEs also expressed difficulties structuring queries to answer questions of interest. Some noted that the system can provide a wealth of information, while others were unsure how to use it: “evaluators don’t know what they’re doing, or they don’t know how to navigate the system of record being DTS.”6 One SME described their views on how the current structure and functionality of DTS could be improved:

I think that there is a way that we could utilize DTS more efficiently. PTBs are sort of the cornerstone of a responder’s experience in DTS. It would be wonderful if—from the cadre side—there was a module in DTS that allowed us to go through that process of creating a workflow to request a modification to a PTB and then that workflow be tracked just like all the other workflows are tracked in DTS. As opposed to the very inefficient [current] process of sending emails back and forth and asking someone to do something. If we were able to put it in a system like DTS where FQS is really housed, it would enable us to track when a request is made, the length of time [the request is outstanding], and who’s responsible for it.7

DTS is largely viewed by those we interviewed as a support element,8 with a SME noting that DTS is useful as “the system of record for whether or not an FQS participant student receives credit for [passed] a course.”9 The same SME commented that some data can be used for tracking purposes because DTS is “where we store information about how much [course] credit is awarded each year or each quarter” but that there are some limitations: “What we do not store is a test score. It’s just simply [whether] a credit was awarded or not. That’s something that I would say is missing from our current system.”10

---

5 Round 2 group interview (#12).
6 Round 2 SME interview (#3).
7 Round 2 group interview (#5).
8 Round 2 SME interview (#4).
9 Round 2 SME interview (#21).
10 Round 2 SME interview (#21).
data in DTS could be addressed by an integrated LMS, which was discussed by SMEs and is described in Chapter 4.

In addition to serving as a storage repository, DTS is instrumental in helping fulfill Federal Coordinating Officers’ requests for IW personnel; however, this capability is not always used as intended. Interviewees reported that while some cadres and deployment regions rely on DTS to select IW for deployment, others often request IW by name. One SME noted that,

I would like the algorithm more if there was some setting [in the algorithm] that when it started to go through candidates it would start to work through higher percentage completion [of PTBs] and then work its way down or allow me to deploy more people by name and actually manage them and their development and making sure that we’re doing right by the disasters a little bit more.11

Other reasons were also provided for not wanting to rely on the DTS algorithm to staff deployments. Trusting and knowing the capability of a specific responder was a common explanation, as was the need to help those closer to PTB completion achieve additional experience, which is also helpful during a disaster:

I like to deploy a candidate who had more experience already, like 75 percent of their PTB completed, because of the nature of the situation and my understanding of what they were going [to be required to do on deployment] as opposed to having somebody [deployed] who just opened that PTB. For the sake of equity and fairness, we try to use the algorithm, but if the algorithm deploys a candidate, it makes no distinction between 0 percent and 80 percent completion.12

Assessing the Effectiveness of the Deployment Tracking System: Findings and Recommendations on What to Measure or Evaluate

SMEs did not discuss metrics for assessing the effectiveness of DTS; however, gaps they identified during interviews suggest some measures and actions that can be used to assess the health of DTS, as described in Table 6.1.

Survey items could address users’ perceptions of DTS effectiveness, understanding of the system, usefulness of the information DTS provides, currency of the information, trust of the algorithm, and effectiveness in selection of personnel for deployment. These questions could be combined into one survey with the items proposed in Chapter 5 to measure satisfaction with QRB decisions. Trust of the algorithm could also be assessed by calculating the

11 Round 2 group interview (#12).
12 Round 2 group interview (#12).
frequency with which users select personnel for deployments by name rather than relying on the algorithm. In addition, FEMA could conduct qualitative analysis of justifications for selecting personnel by name.

**Course of Action Recommendations**

The recommendations below were informed by the interviews and relate to CoAs that FEMA can adopt to improve DTS. They differ from those related to measurement and evaluation.

1. Enlist assistance from the contracting entity that helped develop DTS and create a training aid in the form of a streamlined electronic user guide or FEMA intranet site that explains the DTS algorithm and offers best practices for usage. Publicize its creation to the IW workforce via official FEMA memoranda, emails from the FEMA administrator, or IW-wide emails from supportive senior leaders.

2. Create system modifications to the DTS database to provide more-specific information about the percentage of task completion and/or which PTBs personnel have completed, the effects of which could be measured by periodic end-user surveys (see Table 6.1).
Findings, Recommendations, Observations, and Conclusions

In this chapter we collate findings from the previous chapters, which allows for easy viewing of the findings and recommendations developed for each of the five FQS areas: PTBs, C&Es, training, QRB, and DTS. Table 7.1 presents our findings on FQS strengths and areas for improvement; Table 7.2 presents our recommendations on measuring and evaluating effectiveness across FQS; and Table 7.3 presents the CoA recommendations related to each of the five FQS areas. In addition to collating the recommendations for measures and other CoAs, we provide a user-friendly method for considering how to prioritize efforts based on ease of implementation and time requirements. We then provide observations about gaps and recommendations that cut across the different FQS areas. Because each observation pertains to several FQS areas, we group them in this section rather than in Table 7.3, where recommendations relate to individual FQS areas. The chapter ends with our top-level analysis and concluding comments.

FEMA Qualification System Findings per Area

Overall, FQS has a fair number of strengths, and stakeholders see value in the concept behind FQS and its component parts. However, in practice, the component pieces do not always function as designed. In Table 7.1 we summarize findings regarding the strengths and areas for improvement for the five FQS areas presented in Chapters 2–6.

Gaps in Measuring FEMA Qualification System Effectiveness

This section summarizes the existing and proposed measures of effectiveness for PTBs, C&Es, Training, QRB, and DTS. In brief, we found very few systematic measures and evaluation activities to gauge whether processes in these five components of FQS are implemented effectively and in ways that align with FEMA’s goals. Measures were limited to use of templates and glossaries for HQ review of PTBs and Kirkpatrick Level 1 and 2 evaluations of training effectiveness, though Level 1 assessments are not standardized or systematic and
Level 2 evaluations are sporadic. We did not find measures to evaluate other aspects of PTBs or training or other components of FQS (C&Es, QRBS, and DTS).

For proposed measures—those that are not currently being implemented—we qualitatively evaluated the potential ease of implementation and the potential length of time

<table>
<thead>
<tr>
<th>Core Area</th>
<th>Strength</th>
<th>Area for Improvement</th>
</tr>
</thead>
</table>
| PTB       | • PTBs constructively lay out the tasks required for given position.  
           • PTBs provide a foundation for IW personnel to track career progression.  
           • The PTB revision process involves SMEs to increase efficacy.  
           • There is a lack of trust in the PTB review process; some IW personnel see it as unreliable.  
           • PTBs sometimes omit key information, such as the number of required endorsements.  
           • The application of indicators is not always consistent with FQS Handbook guidance.  
           • Some cadres and C&Es feel pressure to report high PTB completion rates.  
           • Some cadres lack formal PTB review processes.  
           • Training content does not always match PTB requirements. |
| C&E program | • C&Es are designed to be and are seen by many as vital to the training efforts of IW personnel.  
             • There is a wealth of experience brought by C&Es.  
             • There is a shortage of C&Es (lack of personnel to sign off on open PTBs and long time lags in PTB approvals).  
             • C&Es are overburdened by a plethora of tasks.  
             • C&Es sometimes apply PTB ratings inaccurately.  
             • C&Es seemingly apply inconsistent criteria when signing off on PTBs.  
             • There is too little training for C&Es.  
             • C&Es are pressured to qualify cadre members in the field.  
             • Few if any official means exist to evaluate the performance of C&Es. |
| Training   | • The FQS Branch is able to collaborate with cadres to “map the curriculum” to ensure that training aligns with PTBs.  
             • Training schedules and access to training (especially across cadres) result in insufficient availability of classes, impeding readiness.  
             • Roles about branches’ and sections’ responsibilities are unclear, and a lack of collaboration impedes training development, delivery, and evaluation.  
             • A lack of centralized collection or analysis of Kirkpatrick Level 1 data inhibits the use of findings to make improvements in training across cadres or other entities. |
Findings, Recommendations, Observations, and Conclusions

Table 7.1—Continued

<table>
<thead>
<tr>
<th>Core Area</th>
<th>Strength</th>
<th>Area for Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training, continued</td>
<td>• Attempts to conduct Level 2 assessments have been nascent and inconsistent, preventing systematic and robust evaluation of learning outcomes.</td>
<td>• There is a perception that the QRB holds little value because the final qualification determination lays with cadre CA.</td>
</tr>
<tr>
<td></td>
<td>• Level 3 and Level 4 assessments are not being conducted.</td>
<td>• Some feel that the QRB does not provide actionable information for candidates to improve.</td>
</tr>
<tr>
<td></td>
<td>• The lack of an enterprise LMS or LCMS and fragmented IT systems inhibit training development, delivery, evaluation, reporting, and management.</td>
<td>• Cadres question whether QRB members have correct expertise to judge the proficiency of their candidates.</td>
</tr>
<tr>
<td>QRB</td>
<td>• Cadres recognize the QRB’s potential value as an avenue for providing independent, unbiased candidate evaluations.</td>
<td>• The FQS Guide and QRB Handbook give conflicting guidance on whether deployments are required for candidate positions.</td>
</tr>
<tr>
<td></td>
<td>• The QRB can serve as a checkpoint to ensure that candidates meet all the required benchmarks before going to CA review.</td>
<td>• In practice, QRB recommendations are not always impartial and unbiased.</td>
</tr>
<tr>
<td></td>
<td>• Current QRB leadership is focusing on improving the QRB process to increase its usefulness for the cadres.</td>
<td></td>
</tr>
<tr>
<td>DTS</td>
<td>DTS contains and can extract a variety of potentially useful information.</td>
<td>DTS is difficult to structure queries to answer questions of interest.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DTS captures binary pass/fail course data but does not capture full range of test scores.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Trust in the DTS algorithm is inconsistent and some choose to bypass it altogether.</td>
</tr>
</tbody>
</table>

required. These categorizations are intended to help FEMA prioritize efforts to identify or develop and implement effectiveness measures.

The color-coded quad chart shown in Figure 7.1 categorizes the recommendations for measures and related evaluation activities along two axes: a temporal $x$-axis (time needed for implementation) and an operational $y$-axis (ease of implementation). We consider ease of implementation to include such factors as number of resources (including funding), number of personnel, and specific expertise required; needs for coordination among different branches and sections; regulatory requirements; and ability to acquire needed resources.

Using abbreviated prose, we plotted the recommendations from Table 7.2 in the following four quadrants:

- less difficult to implement and requires less time
- less difficult to implement but requires more time
more difficult to implement but requires less time
more difficult to implement and requires more time.

The quadrants are color-coded, with purple representing recommendations that are both less difficult to implement and require less time (lower-hanging fruit) and blue representing recommendations that are more difficult to implement and require more time.

In Table 7.2, we collate measurement and evaluation recommendations from the five FQS areas profiled in this report (PTB, C&E, training, QRB, and DTS). The table captures the full text of the recommendations, along with temporal and implementation ratings for measures that are not currently being implemented. Measures that are currently being implemented are noted. Both the quad chart and table are designed to provide FEMA a quick and easy way to review the proposed recommendations.

Despite our research team’s deliberations to contextualize and assess the measurement and evaluation recommendations along these two dimensions, we acknowledge that FEMA is the foremost authority on what is or is not in its capacity to achieve and in what time frame. Given this, FEMA can use the results from our categorization effort as a starting point for its own discussions about how to press forward with recommendations and the ease or hurdles they might encounter with each.\(^1\)

**FEMA Qualification System Course of Action Recommendations**

This section summarizes the CoA recommendations provided at the end of Chapters 2–6. Like the measurement and evaluation recommendations, we categorized the CoA recommendations in one of four quadrants indicating the relative ease of implementation (e.g., required funding, personnel, expertise, regulatory requirements) and time required for implementation. These results are shown in Figure 7.2. Table 7.3 presents the full text of the CoAs along with temporal and implementation ratings, grouped by topical area (PTB, C&E, training, QRB, and DTS). Both the quad chart and table are designed to provide FEMA a quick and easy way to review the proposed recommendations.

Like decisions for assessing feasibility of adopting measures to assess FQS, FEMA is in the best position to determine which CoAs to pursue and in what time frame. Given this, FEMA can use the results from our categorization effort as a starting point for its own discus-

\(^{1}\) Later conversations with Integration Branch personnel that occurred at the tail end of this research project suggested that a small subset of the recommendations offered in Figure 7.1 might be concurrently underway, albeit with results still forthcoming. We suggest that FEMA engage in further analysis of the recommendations found in each quadrant of Figure 7.1 and prioritize each based on their status as a near-term, medium-term, or far-term objective.
Findings, Recommendations, Observations, and Conclusions

FIGURE 7.1
Measurement and Evaluation Recommendations by Time and Ease of Implementation

- PTB element rubric
- PTB/training alignment rubric
- C&E shortages
- C&E ratings
- Proficiency checklists for C&E
- Increased cadre involvement in training needs analysis
- Level 1 online evaluations
- Level 1 self-efficacy
- Retention of full Level 2 test scores
- QRB perceptions
- DTS perceptions

- C&E IRR
- Analysis of QRB recommendations

TABLE 7.2
Full List of Measurement and Evaluation Recommendations

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Abbreviation Used in Figure 7.1</th>
<th>Time Needed</th>
<th>Ease of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTBs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continue using tools to improve PTB effectiveness (e.g., SMART goals rubric, task template, glossary) during PTB creation and revision</td>
<td>Currently being implemented</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop a standardized rubric or checklist consisting of required PTB elements to ensure accuracy and completeness</td>
<td>PTB element rubric</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Develop a standardized rubric or checklist that documents when tasks from PTBs and content from training are aligned</td>
<td>PTB/training alignment rubric</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
</tbody>
</table>
### Table 7.2—Continued

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Abbreviation</th>
<th>Time Needed</th>
<th>Ease of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&amp;Es</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine appropriate C&amp;E staffing ratios for the IW as a whole and for incident types (by SMEs); analyze administrative data to determine whether staffing ratios are achieved</td>
<td>C&amp;E staffing ratios</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Analyze potential gaps between qualification status (i.e., successful completion of C&amp;E Qualification Course as recorded in DTS) and role assignments for shortages of C&amp;Es</td>
<td>C&amp;E shortages</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Analyze C&amp;E ratings for common rating errors and for possible differences in ratings by protected class or other factors</td>
<td>C&amp;E ratings</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Analyze interrater reliability data—consistency among different C&amp;E ratings of the same candidates</td>
<td>C&amp;E interrater reliability</td>
<td>More time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Institute supervisor oversight and conduct performance review of C&amp;Es</td>
<td>C&amp;E performance reviews</td>
<td>Less time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Conduct 360-degree feedback for developmental, not administrative, purposes</td>
<td>C&amp;E 360-degree feedback</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Provide clear definitions of proficiency in checklists to evaluate T/C proficiency on PTBs</td>
<td>Proficiency checklists for C&amp;E</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continue analysis for training schedules and identify strategies for increased cadre involvement</td>
<td>Increased cadre involvement in training needs analysis</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>For Level 1 evaluations, create a combination of standardized questions and questions that cadres can customize that are administered online</td>
<td>Level 1 online evaluations</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>For Level 1 evaluations, administer online assessments of pre-training and post-training self-efficacy for task performance</td>
<td>Level 1 self-efficacy</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>For Level 2 evaluations, conduct practical exercises (work samples) that reflect the job requirement to assess behavioral outcomes during training, using well-defined checklists and clear options for rating proficiency</td>
<td>Level 2 practical exercises</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>For Level 2 evaluations, create carefully constructed written tests of substantive knowledge for course concepts</td>
<td>Level 2 written tests</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>For Level 2 evaluations, create pre-tests and post-tests to assess improvement in learning</td>
<td>Level 2 pre-tests and post-tests</td>
<td>More time</td>
<td>More difficult</td>
</tr>
</tbody>
</table>
Findings, Recommendations, Observations, and Conclusions

Cross-Cutting Observations and Recommendations

Analysis from the SME interviews and secondary-source literature helped our HSOAC team identify several recommendations and thematic issues that cut across the five FQS areas.

1. According to interviewees’ accounts, we assess there to be a lack of integration and understanding of what different FQS entities and areas do, who is accountable for what, who has decisionmaking authority, and other areas. For example, curriculum

---

2 Later conversations with Integration Branch personnel that occurred at the tail end of this research project suggested that a small subset of the recommendations offered in Figure 7.2 might be concurrently underway, albeit with results still forthcoming. We suggest that FEMA engage in further analysis of the recommendations found in each quadrant of Figure 7.2 and prioritize each based on their status as a near-term, medium-term, or far-term objective.
developers or C&Es might be aware of the QRB and familiar with how it works, but only at a cursory level, with questions about its relevance and overall functionality. The same might be said about a PTB reviewers' familiarity with DTS or a responder’s understanding of how or whether C&Es are evaluated.

Recommendation: The IW’s training entities (FIWA, TDS, and OCCHCO) should build into their instruction plans a mandatory module explaining the interconnectivity of the different parts of FQS. This module, or instruction, can include transparent information about efficacy problems in FQS—a display of honesty that might increase trust and ease with deployers—followed by detailed plans for improvement.

2. We also observed what appears to be a lack of clear top-down messaging and a lack of institutional transparency about many of the FQS challenges that interviewees identified about issues such as the PTB revision process, incomplete or open PTBs, inconsistent criteria employed by C&Es, mismatches between training curricula and PTB contents, relevance issues with the QRB, and the functionality of the DTS algorithm.
### TABLE 7.3
Full List of Course of Action Recommendations

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Abbreviation Used in Figure 7.2</th>
<th>Time Needed</th>
<th>Ease of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PTBs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect insights on the variety of factors that lead to incomplete and complete PTBs by either creating a short survey to distribute among IW personnel or convening focus groups. Publicize results of these efforts via official FEMA memoranda and commit to mitigation measures.</td>
<td>Address causes of incomplete PTBs</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Develop and codify a standard set of expectations of ownership for SMEs, cadre staff, and HQ staff working on creating new PTBs or reviewing and revising existing PTBs to ensure that the individuals involved understand their own roles and how they relate to the roles and responsibilities of others.</td>
<td>Codify PTB creation/revision roles and responsibilities</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Create and internally circulate through the IW a stand-alone graphic showing the interdependencies among the trainer, C&amp;E, and responder.</td>
<td>Communicate interdependencies in PTB roles</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Through official FEMA memoranda, official emails from the administrator, IW-wide emails from senior leaders, and training curricula, emphasize the need to maintain good communication and integrity in C&amp;E sign-offs for the PTB process to be effective and efficient.</td>
<td>Stress the importance of PTB integrity to IW</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Develop and implement a formalized method to process minor but critical PTB changes that arise between official PTB revision cycles.</td>
<td>Formalize process for minor PTB revisions outside the revision cycle</td>
<td>Less time</td>
<td>More difficult</td>
</tr>
<tr>
<td><strong>C&amp;Es</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Through official FEMA memoranda, official emails from the administrator, IW-wide emails from senior leaders, and training curricula: (1) publish and publicize updated and clear expectations for C&amp;Es and (2) publicize the critical role played by C&amp;Es in maintaining the readiness and functionality of the IW, which might incentivize members of the IW to complete the requisite training and assume the role of C&amp;E.</td>
<td>Incentivize the job of C&amp;E and fill more C&amp;E billets</td>
<td>Less time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Create a mechanism for C&amp;Es in the field to inform training personnel at FIWA about T/C readiness problems, which can be addressed by revisions to curricula. C&amp;Es can be mandated to provide generalized feedback on T/C readiness—to inform approaches to training—via email before redeploying from the field.</td>
<td>Create C&amp;E-to-training feedback mechanism</td>
<td>More time</td>
<td>Less difficult</td>
</tr>
</tbody>
</table>
Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions

Table 7.3—Continued

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Abbreviation Used in Figure 7.2</th>
<th>Time Needed</th>
<th>Ease of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hire more C&amp;Es and expand C&amp;E training to focus on obscure PTB tasks and perform evaluations under challenging circumstances. One of several solutions might be increased role-playing or scenario-based training using real world examples of difficult PTB situations, which might boost the confidence of C&amp;Es to sign off on a broad variety of tasks across cadres.</td>
<td>Hire more C&amp;Es and expand C&amp;E training</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>If none currently exists, establish a grievance/redress process for trainees who feel unfairly evaluated by C&amp;Es or neglected altogether and to address potential biases identified in analysis of C&amp;E ratings, as described in Table 3.1.</td>
<td>Create redress process related to C&amp;E evaluations</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Deploy individuals with a dedicated C&amp;E role and remove deployment-related duties outside their C&amp;E responsibilities. Citing the success of the 2019 pilot program to streamline C&amp;E duties (by solely focusing on coaching and evaluating), codify the new practice and publicize the intended outcomes (improved performance of C&amp;Es).</td>
<td>Make C&amp;Es single-hatted</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Training</td>
<td>Provide training, training aids, and access to test-development experts to assist test developers if analyses of the quality of knowledge tests indicate the need for improvement (for example, if the tests do not discriminate among levels of knowledge or behavior).</td>
<td>Provide training for test developers</td>
<td>Less time</td>
</tr>
<tr>
<td>Develop and implement policy that establishes clear responsibilities and ownership of the various aspects of training—including development, delivery, and evaluation for FEMA stakeholders to improve collaboration and coordination.</td>
<td>Develop policy to clarify training roles</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Investigate options, such as individual, simulation-based training, to maintain skill proficiency during long periods of time between deployments for some personnel.</td>
<td>Prevent skill atrophy</td>
<td>Less time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Adopt an enterprisewide LCMS that maximizes needs as identified in the 2021 Capability Analysis Report (FEMA, 2021). However, given that this is not likely to occur in the near future, we recommend exploring the feasibility of purchasing added features of DTS that allow for data capture, storage, analysis, and dissemination.</td>
<td>Acquire LCMS</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Seek fiscal resources for sufficient FEMA and/or contractor personnel—and personnel with appropriate expertise—to increase course availability and support development, analysis, and use of metrics.</td>
<td>Increase training resources</td>
<td>More time</td>
<td>More difficult</td>
</tr>
</tbody>
</table>
A lack of transparency about issues might have unintended consequences for the IW, such as the erosion of trust and low employee engagement. Jiang and Men eval-

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Abbreviation Used in Figure 7.2</th>
<th>Time Needed</th>
<th>Ease of implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>QRB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conduct analyses to determine the ROI of the QRB. An ROI study should balance considerations of perceptions of utility, cost of QRB member and candidate time, impartiality of recommendations, timelines, and potential conflicts with other FEMA goals.</td>
<td>Assess QRB ROI</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Consider an alternate QRB model that better addresses the technical nuances of each cadre and job position. For example, this could take the form of convening, for each cadre, a technical review panel with specialists specific to that cadre.</td>
<td>Consider new QRB model: technical review panel</td>
<td>More time</td>
<td>More difficult</td>
</tr>
<tr>
<td>Continue implementing changes in the QRB process that focus on providing additional utility to cadres. Current QRB leadership is updating documentation and providing additional services to cadres to encourage participation and deepen cadre investment in the process; however, not all cadres have noticed an increase in utility. Continuing these efforts might extend added value to additional cadres and increase the perceived and actual ROI of the QRB.</td>
<td>Continue focusing on QRB updates that provide additional utility to cadres</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Update the QRB Handbook to reflect that the FQS Guide does not necessarily require a candidate to deploy in a position for which they are seeking to be qualified. Allow the QRB to recommend nonqualification on the basis of no deployments (or require a nondeployment waiver) only for those positions that actually require a deployment.</td>
<td>Loosen QRB restrictions on required deployments</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>DTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enlist assistance from the contracting entity that helped develop DTS and create a training aid in the form of a streamlined electronic user guide, or FEMA intranet site, that explains the DTS algorithm and offers best practices for usage. Publicize its creation to the IW workforce via official FEMA memoranda, emails from the FEMA administrator, or IW-wide emails from supportive senior leaders.</td>
<td>Make DTS training aids for IW</td>
<td>Less time</td>
<td>Less difficult</td>
</tr>
<tr>
<td>Create system modifications to the DTS database to provide more-specific information about the percentage of task completion and/or which PTBs personnel have completed, the effects of which could be measured by periodic end-user surveys.</td>
<td>Create modifications to DTS database</td>
<td>More time</td>
<td>Less difficult</td>
</tr>
</tbody>
</table>
uated factors leading to employee engagement. They describe transparent organizational communication, in addition to authentic leadership, as resulting in better employee engagement and work-life enrichment.4

**Recommendation:** FEMA could benefit from increased transparency about FQS processes by communicating information about the strengths and weaknesses of processes to the IW and plans for addressing shortcomings. Once FEMA implements measures of effectiveness, it can publicize results on an online dashboard so that stakeholders have access to information about the health of FQS processes.5

3. The IW could benefit from an annual online survey of stakeholder satisfaction with the FQS processes discussed throughout this report. Survey fatigue is a relevant phenomenon to guard against when conducting workplace climate assessments.6 To minimize potential survey fatigue and enhance participation, FEMA can publicize the importance, relevance, and benefits of participation.7

**Recommendation:** IW leadership can debate the merits of mandating an annual survey designed to measure stakeholder satisfaction with various FQS processes. However, if they choose to make the survey elective, they can boost participation by enlisting senior leadership to tout the mission-oriented nature of the survey, describing the benefits as increased employee satisfaction, increased work-life balance, and a potential for higher IW readiness.

4. Many interviewees mentioned resource shortages—both staffing and funding—as a general challenge for implementing portions of FQS effectively. Suggestions to expand staffing resources as a solution to problems were mentioned in relation to each of the five FQS areas profiled. Three of several examples are interviewees expressing concern that there are not enough instructors to staff training courses; that increased staff numbers could help C&Es squarely focus on coaching and evaluating (rather

---

3 Jiang and Men, 2017.
5 Dashboards such as these are also sometimes referred to as executive dashboards. Companies and organizations use them to transparently display a variety of information for personnel (e.g., financial trendlines over time, customer satisfaction statistics and trends, employee satisfaction, employee hiring versus turnover, and more.) Dashboards tend to focus on different data based on their categorical type. For example, operations and safety dashboards might display workplace accidents and/or health and will want to include resolutions rather than just findings and shift the focus from charts and numbers to the larger goal of promoting safety best practices. A human resources dashboard might be fitting for FEMA. These are common in larger companies and often used to track employee turnover and retention. Those numbers typically give management insight into whether teams have the right capabilities for their roles and highlight potential problem areas. FEMA can create a variation of this, displaying status updates and information about different parts of FQS.
7 Fass-Holmes, 2022, p. 68.
than performing both assessments and deployment duties); and that cadres need more staff to handle their expanding responsibilities.

**Recommendation:** To support a functioning and effective FQS, FEMA should consider prioritizing resource allocation toward staffing for FQS functions—particularly during periods of large-scale efforts—to affect modernization and change. Conducting an in-depth staffing-needs analysis focusing specifically on FQS support staff (such as trainers, dedicated C&Es, and cadre staff) could help FEMA identify the most-critical personnel gaps and priorities.

### Study Strengths and Limitations

A strength of our study was our ability to conduct many interviews \((n = 48)\) and speak with many individuals \((n = 83)\) from various cadres, offices, and entities in the IW. Most conversations lasted approximately 60 minutes, and, when asked, the majority of participants made themselves available for follow-on conversations, which helped us confirm or refute our conclusions. Another strength was the heightened interaction we had with the study’s sponsor, FEMA’s Integration Branch, and the documents, literature, and survey results provided to us over the course of the study, which broadened our understanding of FQS. Frequent conversations with the Integration Branch also helped orient us to understanding FQS and its stakeholders.

Study limitations mainly pertained to access issues. Specifically, despite a goal to speak with the maximum number of cadres (23), we spoke with eight during our Round 2 interviews. Though informative, a more representative population sample might have yielded a broader array of perspectives. Furthermore, although we were able to speak to a large number of SMEs in HQ and cadre coordinator roles, our study did not interview deployers and responders who do not hold those roles. These individuals could provide potentially unique perspectives around FQS culture, adoption, and field implementation. Finally, analysis of the five FQS areas we examined could have potentially been strengthened with in-person visits to FIWA and/or OCHCCO to perform ethnographic research, but that was out of the scope of this research project.

It was also outside the scope of our project to conduct a formal, expansive literature review of past documents and literature related to FQS, though our study team did review the 26 items listed in Figure 1.3 provided by FEMA. We also reviewed externally sourced documents, such as those from GAO. A deeper dive into more literature might have shone light on a wider variety of areas for improvement in FQS or on improvements that have been

---

8 A counterconclusion relates to the concept of data saturation, which is the point in a research process at which enough data have been collected to draw necessary conclusions, and any further data collection might not produce value-added insights. Proponents of data saturation might point out that small studies reach data saturation more rapidly than larger studies. Adherents also often express a preference for quality insights from collected data rather than focusing on quantity.
made over time. For example, documents providing more information on training, LMCS, the QRB, DTS, and advisory boards might have allowed for a more informed perspective on these processes.

Conclusions

FEMA designed FQS to assist in developing and credentialing FEMA IW personnel, using several interrelated processes to accomplish this objective. We examined five of these sections in depth via document review and interviews of 83 SMEs to better understand the ways in which these processes are working well and can be improved and measured. Accordingly, findings in the form of strengths, weaknesses, and approaches to measurement are provided throughout this report. For the latter, interviewees largely described the absence of consistent and codified efforts to measure many of the effectiveness gaps we identified.

Given this, it is not possible to provide a fully informed judgment on the health of FQS that is backed by supporting data. Although we offer recommendations to help FEMA move in the direction of eventually assessing the health of FQS, the lack of current metrics and data render any ability to currently assess FQS incomplete at best. Despite these challenges, we learned over the course of the study that there is a genuine interest in improving FQS, which speaks to an improvement-oriented culture that can be leveraged in the future to measure the effectiveness of and subsequently improve FQS.

Our research team’s overarching perspective is that FQS continues to hold value for the IW despite the gaps and areas for improvement cited throughout this report. However, as identified by GAO and in this report, improvements are warranted for FQS to better serve the needs of a well-trained and deployable IW.9

---

FEMA Qualification System Data Analysis Process

In this appendix, we provide additional information on the data collection and analysis process we employed to learn about FQS, the focal point of which was the SME interviews. An advantage of this study was our ability to speak with a wide swath of individuals from the IW over a span of seven months. To extract the maximum value from the 48 conversations we held, many of which contained more than one individual (we spoke with a total of 83 people), we used a coding and analysis platform called Dedoose (initially described in Chapter 1) to help us identify relevant themes across conversations.

Round 1 and 2 Interviews

Step 1: Codebook Creation

Dedoose coding software helps with text analysis by identifying, searching for, and categorizing meaningful themes across a broad array of documents. This categorization of themes first requires the documents in question to be uploaded to the software. In the context of our study, these documents consisted of 48 de-identifiable (i.e., anonymized) interview transcripts from FEMA's IW. Round 1 interviews spanned August–October 2022 and Round 2 interviews spanned November 2022–February 2023. Interview transcripts were uploaded to Dedoose, and therein we created a set of thematic categories, or codes (referred to as a codebook).

Table A.1 depicts one of two codebooks (for Round 2) we created and its complement of primary and secondary codes. Dedoose, as a content analysis tool, is often used in conjunction with a hierarchical coding schema, or framework, that is validated through inter-rater reliability. Our codebook consisted of a two-tier framework of primary (i.e., parent) codes and secondary (i.e., child) codes that mapped to macro and micro topical areas of our study. For example, parent codes were the macro-level topics, or themes, which included PTBs; C&E; training; QRB; DTS; LMS; and such variables as strengths; weaknesses; timeliness; quality; metrics, evaluation, and assessments; and others. Child codes were the narrower topics that nested under some parent codes and included such codes as training schedule, learning
<table>
<thead>
<tr>
<th>Parent Code</th>
<th>Child Code</th>
<th>Usage Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTBs</td>
<td>Treat this as a macro-level other code for all things pertaining to PTBs that can't be binned under existing child codes.</td>
<td></td>
</tr>
<tr>
<td>Revisions (process for PTBs)</td>
<td>Any mention of the process by which PTBs are reviewed for accuracy or completeness.</td>
<td></td>
</tr>
<tr>
<td>SMEs</td>
<td>Include statements about cadre-specific SMEs and their role in working with or reviewing the PTB.</td>
<td></td>
</tr>
<tr>
<td>Open or incomplete PTBs</td>
<td>Focus on statements describing these two specific phenomena (e.g., open or incomplete PTBs) and their ramifications.</td>
<td></td>
</tr>
<tr>
<td>C&amp;Es (Coaches and Evaluators)</td>
<td>Treat this as a macro-level other code for all things pertaining to coaches and evaluators that can't be binned under existing child codes.</td>
<td></td>
</tr>
<tr>
<td>Quantity of C&amp;Es</td>
<td>Include statements about the quantity of coaches and evaluators: there being enough of them, or a shortage, for example.</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>Treat this as a macro-level other code for all things pertaining to training entities or classes that can't be binned under existing child codes.</td>
<td></td>
</tr>
<tr>
<td>FIWA, TDS, or OCCHCO</td>
<td>Include all statements about training and FIWA or TDS or OCCHCO; statements about who is conducting/developing/evaluating training.</td>
<td></td>
</tr>
<tr>
<td>Alignment of training and PTBs</td>
<td>Include statements about the presence or lack of alignment between PTB tasks and the training that exists or is available in relation to such tasks. Include statements lamenting a lack of alignment.</td>
<td></td>
</tr>
<tr>
<td>Training schedule</td>
<td>Include all statements about training and FIWA or TDS or OCCHCO; statements about who is conducting/developing/evaluating training.</td>
<td></td>
</tr>
<tr>
<td>Learning objectives</td>
<td>Include any comments about learning objectives being developed, considered, desired, discarded, or something different altogether.</td>
<td></td>
</tr>
<tr>
<td>QRB (Qualification Review Board)</td>
<td>Treat this as a macro-level other code for all things pertaining to the QRB or QRB processes. You'll likely double code with other codes such as strengths, weaknesses, recommendations, or others.</td>
<td></td>
</tr>
<tr>
<td>DTS (Deployment Tracking System)</td>
<td>Treat this as a macro-level other code for all things pertaining to DTS.</td>
<td></td>
</tr>
<tr>
<td>LMS</td>
<td>Treat this as a macro-level other code for all things pertaining to LMS. You'll likely double code with other codes such as strengths, weaknesses, recommendations, or others.</td>
<td></td>
</tr>
<tr>
<td>Strengths</td>
<td>Include statements explicitly identifying things as strengths, or something that is working well. You'll likely double code with other codes such as PTBs, DTS, or others.</td>
<td></td>
</tr>
</tbody>
</table>
### Table A.1—Continued

<table>
<thead>
<tr>
<th>Parent Code</th>
<th>Child Code</th>
<th>Usage Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weaknesses</td>
<td>Include statements explicitly identifying things/processes as weak, or needing improvement. You’ll likely double code with other codes such as PTBs, DTS, or others. Exclude statements about areas or types of improvements needed and instead code as recommendation or double code with a topical code (e.g., [LMS], if the respondent says the current LMS system is un-scoped and needs to be revamped, for example).</td>
<td></td>
</tr>
<tr>
<td>Purpose, description, or process</td>
<td>Include statements about how relevant processes or offices were started, or about their current composition. Examples: descriptions about the PTB process; or about how DTS came about and how it is structured; or how the QRB was started and its current composition. Also include aspirational statements about what something was designed to do (but perhaps never materialized).</td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>Include statements in which respondents comment on the quality (or lack thereof) of a process, position, or something commensurate. Such statements can be neutral in nature. Exclude statements describing select practices, policies, or FQS processes as strong or weak. Instead, code as Strengths or Weaknesses.</td>
<td></td>
</tr>
<tr>
<td>Timeliness</td>
<td>Any mention of timeframe/timeliness being an issue or working well.</td>
<td></td>
</tr>
<tr>
<td>Recommendations</td>
<td>Include any direct or indirect recommendation offered by respondent/s. Keep a keen eye out for indirect statements in which respondents ruminate about an ideal state of affairs (or an ideal policy or process) without clearly labeling it a “recommendation.”</td>
<td></td>
</tr>
<tr>
<td>Metrics, evaluation, assessment</td>
<td>Include statements about metrics and measurements that currently exist. Also include statements where respondent/s hypothesize about future metrics or lament the current lack of metrics. Include statements using similar terms such as “evaluation” or “assessment.”</td>
<td></td>
</tr>
<tr>
<td>Overburdened/too many hats</td>
<td>Include statements about cadre members, or members of other offices (i.e., PTB or QRB) asked to do too many tasks, resulting in feelings of being overwhelmed. For examples, C&amp;E who deploy and must both evaluate cadre members and perform their deployment tasks. Or cadre members who also serve on the QRB and can’t do both satisfactorily.</td>
<td></td>
</tr>
<tr>
<td>Zinger quotes</td>
<td>Scan interviews for short, catchy but relevant quotes</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>Include all statements that seem important but don’t fit in our existing categories. Err on the side of binning such statements here rather than ignoring altogether. (Remember: when in doubt . . . code)</td>
<td></td>
</tr>
</tbody>
</table>
Identifying Gaps and Areas for Improvement in the FEMA Qualification System for Incident Workforce Positions

objectives, and alignment of training and PTBs (nested under the parent code training) and revisions, SMEs, and open or incomplete PTBs (nested under the parent code PTB).

Step 2: Inter-Rater Reliability

The extent of agreement among data collectors is called inter-rater reliability, or IRR. Though Dedoose helps research teams build and maintain IRR, which ensures that researchers share a common understanding of codes and apply these codes to text in a consistent and accurate manner, we instead opted for a self-designed IRR process that required less time but still ensured consistency and unanimity between coders.

This began with a thorough review of the codebook to ensure that analysts had a strong working knowledge of the codes and their associated definitions. Coding best practices, and common mistakes to guard against, were discussed in detail. Then, all five analysts individually coded a single document on their own and reconvened thereafter to compare coding decisions and adjudicate disagreements over the application of codes. Coding application error is common in the beginning stages of group coding when multiple people are parsing the same document. As is common with IRR exercises, we found instances of inconsistent application of codes. Two consensus-building meetings that debated the codebook and application choices and patterns helped ensure uniformity going forward.

Using a spreadsheet that listed completed interview transcripts, analysts self-nominated to code documents. As the coding process unfolded and continued, questions about code application or other challenges were adjudicated across the team to help ensure accuracy.

Step 3: From Coding to Analysis

After achieving consensus on the meaning and application of codes in Dedoose, our coders reviewed every sentence and paragraph of the 48 imported transcripts and binned or categorized the content based on the codes or themes. This resulted in a total of 365 interview excerpts that were tagged with codes such as overburdened/too many hats, timeliness, or recommendations. Overall, a total of 1,086 codes were applied, with many excerpts being tagged with multiple codes.\(^1\) After content from every document was binned or coded, the research team engaged in the analytical process of deriving meaning from the 365 coded excerpts.\(^2\)

---

\(^1\) This process of double, triple, or quadruple coding is quite common and entails applying several codes to a single excerpt, sentence, or set of sentences. For example, if an interviewee comment included a recommendation on how to improve the alignment of training and PTBs for the purpose of reducing stress among SMEs, that excerpt might have four codes applied: weaknesses, alignment of training and PTB, SMEs, and recommendations.

\(^2\) We only feature our Round 2 codebook in this appendix, though we also used a similar schema in Dedoose schema Round 1 that also included parent codes \(n = 11\), child codes \(n = 34\), and related inclusion rules.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&amp;Es</td>
<td>coaches and evaluators</td>
</tr>
<tr>
<td>CA</td>
<td>Certifying Authority</td>
</tr>
<tr>
<td>CoA</td>
<td>course of action</td>
</tr>
<tr>
<td>COVID-19</td>
<td>coronavirus disease 2019</td>
</tr>
<tr>
<td>DTS</td>
<td>Deployment Tracking System</td>
</tr>
<tr>
<td>FD</td>
<td>Federal Directive</td>
</tr>
<tr>
<td>FEMA</td>
<td>Federal Emergency Management Agency</td>
</tr>
<tr>
<td>FFRDC</td>
<td>federally funded research and development center</td>
</tr>
<tr>
<td>FILS</td>
<td>FEMA Integrated Learning System</td>
</tr>
<tr>
<td>FIWA</td>
<td>FEMA Incident Workforce Academy</td>
</tr>
<tr>
<td>FQS</td>
<td>FEMA Qualification System</td>
</tr>
<tr>
<td>GFI</td>
<td>government-furnished information</td>
</tr>
<tr>
<td>HQ</td>
<td>headquarters</td>
</tr>
<tr>
<td>HSOAC</td>
<td>Homeland Security Operational Analysis Center</td>
</tr>
<tr>
<td>IRR</td>
<td>inter-rater reliability</td>
</tr>
<tr>
<td>IT</td>
<td>information technology</td>
</tr>
<tr>
<td>IW</td>
<td>incident workforce</td>
</tr>
<tr>
<td>LCMS</td>
<td>learning content management system</td>
</tr>
<tr>
<td>LMS</td>
<td>learning management systems</td>
</tr>
<tr>
<td>OCCHCO</td>
<td>Office of the Chief Component Human Capital Officer</td>
</tr>
<tr>
<td>PTB</td>
<td>Position Task Book</td>
</tr>
<tr>
<td>QRB</td>
<td>Qualification Review Board</td>
</tr>
<tr>
<td>ROI</td>
<td>return on investment</td>
</tr>
<tr>
<td>SMART</td>
<td>specific, measurable, achievable, relevant, and time-bound</td>
</tr>
<tr>
<td>SME</td>
<td>subject-matter expert</td>
</tr>
<tr>
<td>T/C</td>
<td>trainee/candidate</td>
</tr>
<tr>
<td>TDS</td>
<td>Training Development Section</td>
</tr>
<tr>
<td>WDD</td>
<td>Workplace Development Division</td>
</tr>
</tbody>
</table>
References


Federal Emergency Management Agency, “FEMA Disaster Levels and Incident Support Staff Activation,” webpage, undated. As of July 12, 2023: https://emilms.fema.gov/is_0822/groups/64.html


FEMA—See Federal Emergency Management Agency.

GAO—See Government Accountability Office.


To fulfill its mission objectives, the Federal Emergency Management Agency (FEMA) requires a well-trained workforce. The Post-Katrina Emergency Management Reform Act of 2006 mandated that FEMA develop a personnel credentialing process for incident workforce personnel responding to disaster situations. FEMA, in turn, created the FEMA Qualification System (FQS), a credentialing process that focuses on standardizing personnel abilities through training and applied experience.

Following concerns raised about the efficiency and effectiveness of FQS to provide the needed structure to qualify individuals for their assigned positions, FEMA’s Integration Branch asked the authors to identify existing FQS gaps and areas for improvement and examine ways to measure and monitor the effectiveness of FQS in the future. This report details the authors’ approach to addressing both issues and the related findings so that FEMA can ensure the development of employees who provide vital disaster recovery services to the United States and its people.