

COGPLEXITI · SYSTEMS & DECISION ANALYSIS

CIMT SKILLS DEVELOPMENT FRAMEWORK

A Five-Level Path from
Qualified to Expert for Complex
Incident Management Team
Personnel

JAYSON COIL

COGPLEXITI · SYSTEMS & DECISION ANALYSIS
FRAMEWORK

CIMT Skills Development Framework

*A five-level path from qualified to expert for Complex Incident Management
Team personnel*

For expert-panel review

Generalized from the DIVS framework (v1). v3 added figures, Appendices B–E, and references; v4 reconciles citations to verified sources only; v5 adds the pilot instrument set (Appendix F), the evaluator calibration protocol (Appendix G), and the Command scope statement in Section 9.

COPYRIGHT AND LICENSE

© 2026 Jayson Coil. All rights reserved.

ISBN 979-8-9970824-4-4

Published by Cogplexiti LLC.

No part of this publication may be reproduced, distributed, or transmitted in any form or by any means without the prior written permission of the publisher, except for brief quotations used in reviews and other uses permitted by copyright law.

Use for Artificial Intelligence and Machine Learning. This work may not be used to train or develop artificial intelligence or machine learning systems, or be included in any dataset used for such purposes, without the prior written permission of the publisher.

Permissions. Direct requests to Cogplexiti LLC at permissions@cogplexiti.com.

EDITION INFORMATION

July 2026. Published by Cogplexiti LLC, Flagstaff, Arizona · jayson@cogplexiti.com · www.cogplexiti.com

Part of the Cogplexiti Analysis line (Systems & Decision Analysis), alongside *The System Tier*, *The Incident Tier*, *Sustaining Decision Performance*, and *The Reveal*.

DISCLAIMER

The views, frameworks, and recommendations expressed in this publication are those of the author and do not necessarily reflect the official policy or position of the Sedona Fire District, Southwest Incident Management Team 1, the National Wildfire Coordinating Group, the U.S. Forest Service, or any other agency or organization. This document is not doctrine. It is a professional analysis informed by operational experience, research, and study.

Appendices B through E and the Appendix F–G instruments are working drafts for expert-panel review; nothing in them is a requirement until it passes the pilot logic in Section 9.

CONTENTS

1 · Purpose	5
2 · The Five Levels	6
3 · What Each Level Means	7
4 · The Second Skill: Checking Your Own Thinking	9
5 · Building the Skills List for Each Position	10
6 · Skill Areas Shared by Every Position	11
7 · How Progress Is Evaluated	12
7.1 The rating tool	12
7.2 Three sources of evidence – never just one	12
7.3 Rules for every gate	13
7.4 Schedule	14
8 · What Can Go Wrong	14
9 · Rollout	15
10 · Corpus Anchors	16
Appendix Template – What Every Appendix Must Contain	18
Appendix A – Operations Section: Division/Group Supervisor	19
A.1 Level 1 – Newly Qualified: Working Through Structure	19
A.2 Level 2 – Developing: Learning What Matters	20
A.3 Level 3 – Competent: Deliberate Division Management	21
A.4 Level 4 – Proficient: Reading the Whole Division	22
A.5 Level 5 – Expert: Trusted Judgment, Checked	22
A.6 Evaluator Pool and Empty Cells	23
Appendix B – Planning Section: Situation Unit Leader	24
B.1 Level 1 – Newly Qualified: Working Through Structure	24
B.2 Level 2 – Developing: Learning What Matters	25
B.3 Level 3 – Competent: Deliberate Picture Management	26
B.4 Level 4 – Proficient: Reading the Whole Information System	26
B.5 Level 5 – Expert: Trusted Picture, Checked	27
B.6 Evaluator Pool and Empty Cells	28
Appendix C – Logistics Section: Logistics Section Chief	29
C.1 Level 1 – Newly Qualified: Working Through Structure	29
C.2 Level 2 – Developing: Learning What Matters	30
C.3 Level 3 – Competent: Deliberate Support Management	30

C.4 Level 4 — Proficient: Reading the Demand Curve.....	31
C.5 Level 5 — Expert: Trusted Forecast, Checked	32
C.6 Evaluator Pool and Empty Cells.....	32
Appendix D — Finance/Administration Section: Finance/Admin Section Chief	33
D.1 Level 1 — Newly Qualified: Working Through Structure	33
D.2 Level 2 — Developing: Learning What Matters.....	34
D.3 Level 3 — Competent: Deliberate Fiscal Management	34
D.4 Level 4 — Proficient: Reading the Fiscal Trajectory	35
D.5 Level 5 — Expert: Trusted Numbers, Checked	36
D.6 Evaluator Pool and Empty Cells.....	36
Appendix E — Command Staff: Safety Officer	37
E.1 Level 1 — Newly Qualified: Working Through Structure.....	37
E.2 Level 2 — Developing: Learning What Matters	38
E.3 Level 3 — Competent: Deliberate Risk Management.....	38
E.4 Level 4 — Proficient: Reading the Incident’s Risk Trajectory	39
E.5 Level 5 — Expert: Trusted Judgment, Modeled	40
E.6 Evaluator Pool and Empty Cells.....	40
Appendix F — Pilot Instrument Set.....	41
F.1 — DIVS Rating Sheets	41
F.2 — Simulation Score Sheet	45
F.3 — Exposure Log and Difficulty Coding Scheme	46
F.4 — Structured Reflection Template	48
F.5 — Gate Record	49
Appendix G — Evaluator Calibration Protocol	49
G.1 — Purpose and Standard	49
G.2 — Session Design	50
G.3 — Cadence and Pool Rules.....	50
G.4 — Records.....	50
References.....	51

1 • Purpose

The qualification system tells us when a person is ready to start a job. It tells us nothing after that. A position task book is a one-time check: an evaluator watches the person do the required tasks, signs the book, and the person is qualified. After the signature, the system assumes the person will keep getting better just by doing the job.

That assumption fails in wildland fire. Fire assignments do not teach lessons automatically. A person can get a good outcome for the wrong reasons and never know it. Experience without structure builds confidence faster than it builds skill. And the people most likely to overrate their own ability are the ones who have never faced conditions hard enough to show them their limits.

Right now, the system cannot answer three basic questions about any qualified member of a Complex Incident Management Team (CIMT):

- Where does this person sit on the path from newly qualified to expert?
- What should this person work on next?
- Is this person's ten years of experience real growth — or one year repeated ten times?

This framework is built to answer those questions. It gives every CIMT position a clear five-level path from newly qualified to expert. It tells each member exactly what to work on to reach the next level. And it checks progress by watching what people actually do — not by counting years or collecting signatures.

Three parts. This base document explains the system: the five levels, the method for building the skills list for each position, and the rules for evaluating progress. The appendices apply the system to each functional area of the team — Operations, Planning, Logistics, Finance/Administration, and Command Staff. Appendix A (the Division/Group Supervisor, in the Operations Section) is complete and serves as the worked example; every appendix matches its structure. Appendices B through E are drafts, built for expert-panel review under the Section 5 method — nothing in them is a requirement until it passes the pilot logic in Section 9. Appendices F and G carry the instruments the pilot runs on — the rating sheets, simulation score sheet, exposure log and difficulty codes, reflection template, gate record, and the evaluator calibration protocol. The instruments measure; they do not yet require.

Intent. This is a development tool, not a promotion board. Its job is to help people grow and to make that growth visible. Each level gives the member a short list of things to work on — four or five, never more. A member at any level should be able to say from memory what they are working on. The evaluation rules exist to keep those lists honest.

Where this comes from. This framework does not invent new ideas. It applies ideas already tested and documented in the Cogplexiti body of work: the evaluation chapters of the *Operational Leader's Field Guide*, the meta-recognition paper *The Second Skill*, the certification rules in *The Incident Tier*, and the system-level case made in *Sustaining Decision Performance*. Section 10 lists where each piece comes from, so the framework can be checked against its sources.

2 · The Five Levels

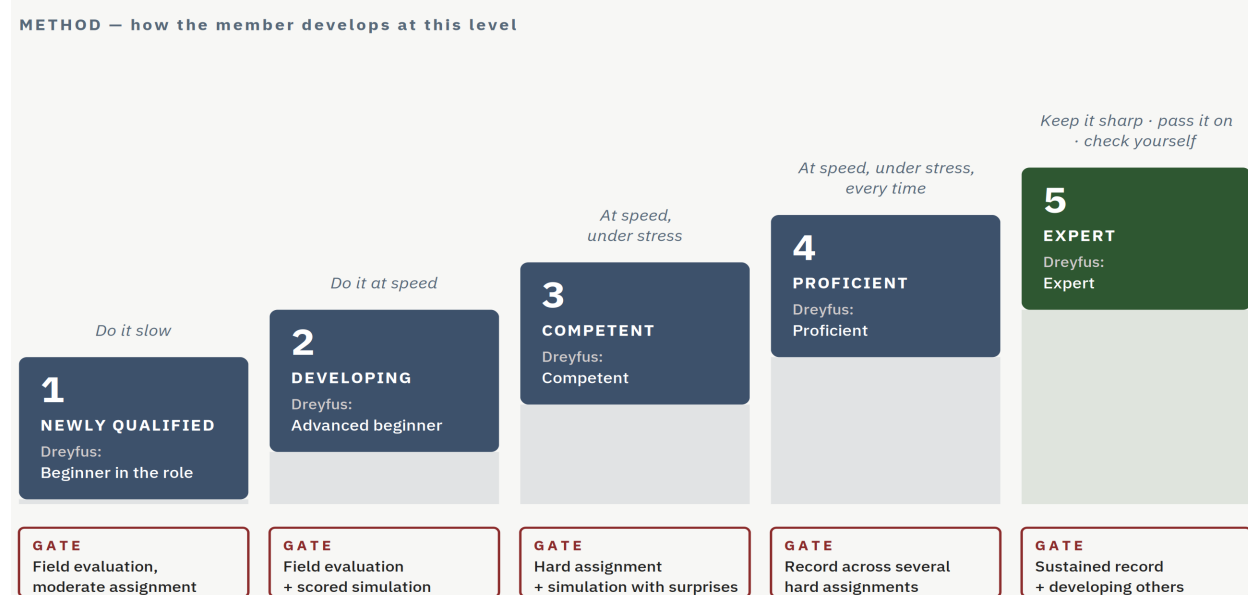
The framework uses five levels. They come from the Dreyfus model of skill acquisition, a well-known description of how people move from beginner to expert (Dreyfus and Dreyfus 1986). The Dreyfus model matters because it describes how a person handles a situation — not how many hours they have logged. The difference between levels is the quality of the thinking, not the amount of time.

One rule governs movement. It is called the developmental progression: do it slow, then do it at speed, then do it at speed under stress, then do it at speed under stress consistently. Each step raises the load, not just the repetitions. A person moves up when they demonstrate performance under the right load — never because time has passed, and never because paperwork says so.

FIGURE 1

The five-level path

Each level pairs a development method with a gate. The method says how the member practices; the gate says what evidence advances them. Time in grade appears nowhere.



Levels come from the Dreyfus model — the difference between them is the quality of the thinking, not time in grade.

The difference between levels is the quality of the thinking, not the amount of time.

Where a new person starts. A newly qualified member is a beginner in that position, not a beginner in fire. Most people arrive with years of experience, and that experience carries over. What does not carry over is the position's decision territory — the scale of the choices, the authority, and the risk that come with the new job. In that territory, the member starts the progression fresh, working through structure and rules the way beginners do in any field. Experienced people re-enter at the early levels when the

decision territory changes, and they move through those levels faster than true beginners — but they do move through them. The task book records entry into the territory. It says nothing about position on the path through it.

Level	Dreyfus stage	How they develop	Thinking skill to build	How the gate is checked
1 — Newly Qualified	Beginner (in the role)	Do it slow	Say your assumptions out loud	Field evaluation on a moderate assignment
2 — Developing	Advanced beginner	Do it at speed	The quick test: when to stop and double-check	Field evaluation + scored simulation
3 — Competent	Competent	At speed, under stress	Audit your assumptions; count what you have explained away	Field evaluation on a hard assignment + simulation with surprises
4 — Proficient	Proficient	At speed, under stress, every time	Change your read when the facts turn against it	Record across several hard assignments + simulation
5 — Expert	Expert	Keep it sharp; pass it on; check yourself	Think out loud so others learn the habit	Sustained record + developing others

3 · What Each Level Means

The descriptions below are position-neutral. Each appendix rewrites them in the operational language of its own positions. The focus areas listed here are themes; the appendices turn them into specific, observable behaviors.

Level 1 — Newly Qualified: Working Through Structure

The member runs the job by the book. Briefings follow the standard format, rules are applied as written, and the plan is executed as it was given. The member sees what is happening but does not yet know which details matter most. When the situation stops matching the plan, the member asks for help early — and at this level, asking early is the right move, not a weakness.

Focus areas build habits: communicating completely, building a full picture of the situation, executing the plan while being able to explain the reason behind it, saying assumptions out loud, and asking for help at the right time. Method: do it slow — assignments of moderate difficulty, close supervision, and a review before and after each shift. Gate: a field evaluation rated “meets” on every Level 1 behavior. Time in the job proves nothing by itself.

Level 2 — Developing: Learning What Matters

The member starts to tell important cues from background noise and recognizes familiar situation types. But the member still treats too many things as equally important, and is still at risk of plan lock — continuing to run a plan after the conditions it was built on

have changed. The most common Level 2 failure is not missing the warning sign. It is seeing the sign and accepting an easy explanation for it, because the explanation lets the current plan stay alive.

Focus areas: building decision triggers and defending the math behind them, naming what each choice buys and what it costs, noticing when the job is outgrowing the plan, managing resources against a stated purpose, and learning the quick test (Section 4). Method: do it at speed — full-tempo assignments, with simulation beginning as a scored practice space. Gate: field evaluation plus scored simulation.

Level 3 — Competent: Deliberate Management

The member sees the assignment as a system that runs across multiple days, not one shift at a time. Planning is deliberate. Backup plans are built ahead of time instead of improvised. Coordination with neighboring units is managed, not just reacted to. The work is effective, but it takes real effort — and that effort is the defining mark of this level. Level 3 is the floor for independent assignment to difficult work, and it is where most careers will sit for most of their length. That is normal, not a failure.

Focus areas: stating the reasoning behind each big choice before acting, building backup plans with named triggers, auditing your own assumptions at set points in the day, communicating clearly under pressure in both directions, and managing the boundaries with the units around you. Method: at speed, under stress — assignments hard enough to actually test these skills, because skill shown only on easy assignments is assumed, not proven. Gate: field evaluation on a genuinely hard assignment, plus simulation with built-in surprises, scored on one behavior above all: when the facts turn against the member's read of the situation, do they change the read or defend it?

Level 4 — Proficient: Reading the Whole Picture

The member sees immediately what matters and what is noise, anticipates instead of reacting, and makes most decisions with little effort — which frees mental capacity for exactly the moments that demand it. The risk at this level is the mirror image of the strength: fast pattern matching applied to a situation that has quietly stopped matching the pattern. The Level 4 focus areas exist mostly to guard against that.

Focus areas: noticing when the situation is genuinely new and slowing down; holding more than one read of the situation open instead of settling too early; dropping a working read when the evidence turns against it — including a read the member built and briefed personally; shaping what information moves up the chain so warning arrives while it is still cheap to act on; and running trainees deliberately rather than as passengers. Method: at speed, under stress, consistently. Gate: the gate shifts from one evaluated assignment to a record built across several — sustained “meets or exceeds,” demonstrated read-changing behavior in field and simulation, and evidence of trainee development.

Level 5 — Expert: Trusted Judgment, Checked

The member runs on deep pattern recognition, saving deliberate analysis for genuinely new situations and for the moments the quick test flags. The self-check runs as habit — and the expert runs it out loud, so others learn the habit by watching. Expertise fades

without use. Skills that are not exercised against truly hard conditions decay quietly, and every routine assignment that goes fine hides the decay. So Level 5 carries maintenance duties, not just status.

Focus areas: developing others as the primary output; narrating your own thinking in public; honestly naming the areas where your expertise has not been tested by current conditions and operating there the way a Level 3 would; and keeping exposure at the difficulty your skills were built for — with simulation filling in when the season does not provide it, and the gap recorded rather than ignored.

The precedent for evaluating experts. Airline pilots take proficiency checks every six to twelve months. Nuclear plant operators requalify about every two years, with simulator tests. Military commanders face regular evaluation and comparative selection. Those fields share fire's features — shared authority, outcomes tangled with outside causes, judgment-heavy work — and they built recurring evaluation anyway (*Sustaining Decision Performance*).

4 · The Second Skill: Checking Your Own Thinking

Experienced people decide by pattern matching. They see a situation, recognize it as a familiar type, pull up the response that type calls for, run a quick mental check, and act. That is what expertise is, and the fire environment demands it. That is the first skill, and the qualification system trains it well.

The second skill is checking your own pattern match while it is still running. The system assumes this skill shows up on its own with experience. Research says it does not (Cohen, Freeman, and Thompson 1995). It is a separate skill — and the good news is that it can be taught in hours, not years, and the effect can be measured. This framework builds it in at every level, scaled to the level. The tools are simple:

- The quick test. Know when to stop and double-check. The rule: when the stakes are high, when there is time, and when the situation is unfamiliar. When stakes are low or time is gone, trust the pattern and act.
- The crystal ball drill. Take the read you are most sure of. Pretend a perfect source just told you it is wrong. Your job is to explain how that could be — then explain it a second way, and a third. Every explanation you generate is an assumption you were carrying without knowing it. Most of the time your read survives the drill. But now your confidence is earned instead of assumed, and you hold two or three named risks you did not have before.
- The conflict count. Any solid read can absorb one surprising cue — there is usually a reasonable explanation. The danger is not the explaining. The danger is losing track of how much explaining your read now needs. So ask one question at set points during the day: how many things have I explained away today? Treat the second one as a signal, not a coincidence.
- Changing the read. When the count gets high, or the drill turns up something serious, build a new story of what is happening instead of patching the old one. This

is the hardest move — especially when the old story is one you built and briefed yourself — and it is exactly the behavior the upper-level gates score.

5 · Building the Skills List for Each Position

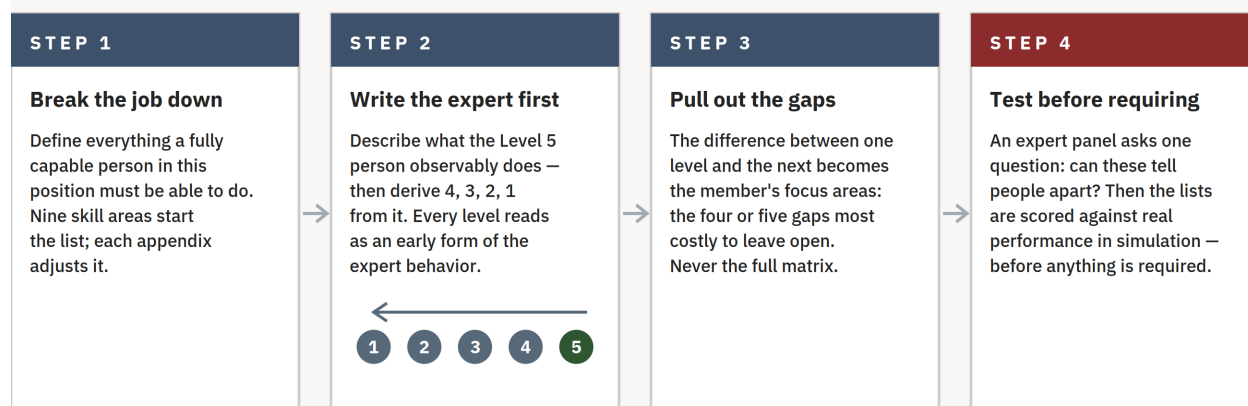
The levels above are the skeleton. The specific behaviors that fill them in are different for every position — a Division Supervisor and a Logistics Section Chief do not need the same skills. Every appendix builds its behavior lists using the same four steps, in this order. The order matters: define what the job requires first, then derive the levels from it. Working the other direction fills the levels with whatever was easy to name.

FIGURE 2

Build backward

The four-step build runs in order for every position. The expert behavior is written first; every lower level is derived from it, so each skill area tells the story of one skill maturing.

THE FOUR-STEP BUILD — RUN IN ORDER FOR EVERY POSITION



*The order matters: define what the job requires first, then derive the levels from it.
Working the other direction fills the levels with whatever was easy to name.*

Requirements follow measurement, always.

Step 1 — Break the job down

Define everything a fully capable person in this position must be able to do, at any level. A workable starting list of skill areas — each appendix adjusts it:

- Situation assessment — building the picture of what was, what is, and what is coming, plus the warning signs that would say the picture is wrong
- Communication — complete messages, checking for double meanings, catching softened warnings
- Decision-making — pattern matching, and the checks on it (Section 4)
- Risk management — safety systems, decision triggers, and accepting risk deliberately

- Resource and workload management
- Coordination — the units beside you, above you, and the ones you depend on
- Backup planning and adaptation
- Self-management and cognitive readiness — fatigue, mental load, and the safeguards that catch what a tired person cannot catch in themselves
- Developing others — weighted more heavily at Levels 4 and 5

One note on the eighth area. Fatigue damages the exact self-judgment a person would use to notice they are fatigued. A worn-down member feels functional — and is wrong. So this area is scored by watching whether the member uses the safeguards: go/no-go checklists, a second person's concurrence on critical calls, and keeping big decisions away from the low points of the body clock. It is never scored by asking the member how they feel.

Step 2 — Write the expert behavior first, then work backward

For each skill area, describe what the Level 5 person observably does. Then describe the same behavior at Level 4, then 3, then 2, then 1. Working backward keeps the levels connected: every Level 2 behavior should read as an early form of the Level 5 behavior, so each skill area tells the story of one skill maturing — not five unrelated requirements. It also keeps the lists honest. Some areas (developing others, for example) genuinely start at Level 3 or 4. The list should say so instead of inventing filler.

Step 3 — Pull out the differences as focus areas

The gap between one level and the next, in each skill area, is written as something an evaluator could actually see in the field: what the Level 3 person does that the Level 2 person does not. The focus areas handed to each member are the four or five most important gaps to the next level — chosen by what it would cost to leave them open on an assignment, not by what is easiest to teach. The member is never handed the full matrix. They are handed their gaps.

Step 4 — Test before requiring

A panel of experienced people who have held the position reviews the behavior lists with one question: can these actually tell people apart? A behavior every qualified person already shows, or one that no one could observe in the field, gets rewritten or cut. The surviving lists are then tried in existing simulation courses and scored against real performance — before any of it becomes a requirement. Requirements follow measurement, always. The cost of a pilot is the price of not scaling a tool that does not work.

6 · Skill Areas Shared by Every Position

Four skill areas appear in every appendix: communication, situation assessment, cognitive readiness, and developing others. They are defined once, here, so that a change made here reaches every appendix at the same time — instead of five appendices slowly drifting apart.

- **Communication** uses the team’s standard formats: complete transmissions built on the SAPPER structure, plain-language checks for phrases that could mean two things, and the discipline of un-softening important warnings — in both directions, hearing them and sending them.
- **Situation assessment** builds the picture in three tenses — what was, what is, what is coming — and always attaches a fourth piece: what would tell you the picture is wrong.
- **Cognitive readiness** is scored as safeguards in use, as described in Section 5 — never as self-report.
- **Developing others** means structured exposure, reasoning made visible, and practice corrected — running trainees deliberately, not carrying them as passengers.

Each appendix writes position-specific behaviors for these areas. The definitions and the level structure come from here.

7 · How Progress Is Evaluated

7.1 The rating tool

Every skill area at every level is rated on a three-point scale: does not meet, meets, exceeds. Each rating is defined by specific behaviors an evaluator can watch for — not by the evaluator’s general impression. This approach is called a behaviorally anchored rating scale, and it is the standard for fair, consistent, defensible evaluation (Smith and Kendall 1963). It changes the evaluator’s question from “how did I feel about the performance?” to “which of these specific behaviors did I see?” A member’s current level is wherever they rate “meets” across all the skill areas that apply. Their focus areas are the gaps to the next level. The instruments themselves — rating sheets, simulation score sheet, exposure log, reflection template, and gate record — are in Appendix F.

7.2 Three sources of evidence — never just one

- **Field evaluation on hard-enough assignments.** The main evidence for Levels 1–3, and part of every gate. The assignment must be hard enough to test the level being certified — a Level 3 gate cannot be passed on an assignment that never stressed the Level 3 behaviors. Every assignment gets a difficulty code, kept in a simple exposure log (the coding scheme is Appendix F.3), so the record is honest about what the member has actually been tested against.
- **Scored simulation.** The place where the thinking skills get measured, because field observation only catches them by luck. Simulation exercises already exist; what is added is a score sheet for observable behaviors — assumptions said out loud per scenario, warning signs acknowledged versus explained away, reads changed when the scenario turns — and a standard for each level. Simulation is also the fairness

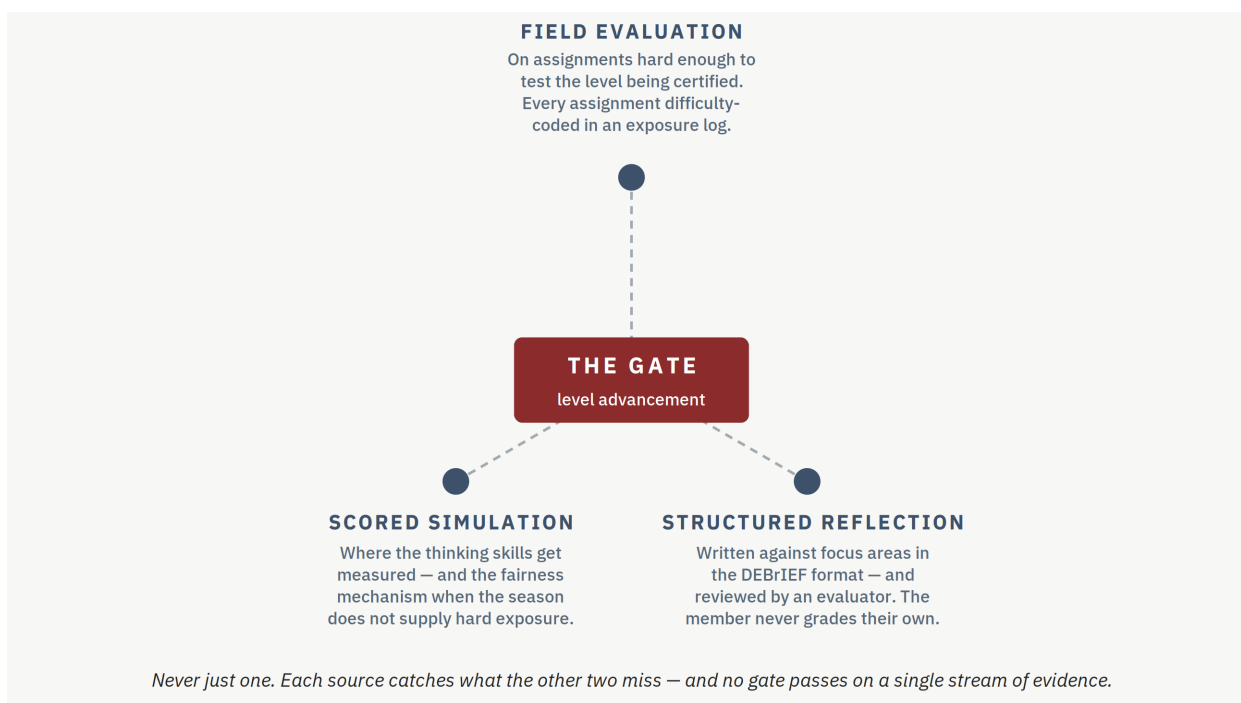
mechanism: it supplies the hard exposure that a mild season or a quiet part of the country does not.

- Structured reflection, reviewed by someone else. After each assignment, the member writes a short record against their focus areas: what they tried, what happened, and what they now believe and why. It follows the DEBrIEF format the team already uses, and an evaluator reviews it — the member never grades their own. The research here is strong: structured debriefs measurably improve performance (Tannenbaum and Cerasoli 2013). Self-review without structure just confirms what the person already believed.

FIGURE 3

Three sources of evidence

Every gate draws on all three streams. Field evaluation catches performance under real load; simulation catches the thinking skills the field only shows by luck; reviewed reflection catches what the member is learning from both.



Never just one — no gate passes on a single stream of evidence.

7.3 Rules for every gate

- Performance, never a paperwork transition. There is no administrative path between levels — no advancing on coursework alone, no equivalency memos. A label stays connected to a capability only if the label is earned by demonstrated

performance under real load. Any shortcut around that is a relabeling, whatever the memo calls it.

- The evaluator standard. Evaluators hold the level being certified, or higher, and attend regular calibration sessions where they rate the same recorded scenarios and work out their differences (the calibration protocol is Appendix G). Two more rules protect the system. An evaluation done by someone who has not genuinely operated at the level being assessed is marked as such in the record — not treated as equal. And honest evaluations must be professionally safe to give: an evaluator under pressure from promotion timelines or home units will drift toward friendly ratings, and the system will get the ratings it rewards instead of the ratings it needs.
- Behavior, not forms. Every thinking skill is scored as live behavior in simulation or the field — never as documentation. The moment the assumption check becomes a form to fill out, the drill has turned into compliance, and compliance is being scored while the skill goes unmeasured.
- A plan attached to every decision. Every gate result — pass or hold — comes with a named set of focus areas and an exposure plan for the next cycle. A hold with a plan is development. A hold without one is just a rejection.

7.4 Schedule

Level review happens at the end of each season, or after any evaluated assignment at or above the member's target difficulty. Focus areas are set at the review, carried through the season, and referenced in every post-assignment reflection and AAR the member takes part in. The schedule keeps the framework present in the work instead of becoming annual paperwork. A member should always be able to answer three questions: what is my level, what are my gaps, and what evidence does my next gate need?

8 · What Can Go Wrong

One standing objection and six failure modes. Each has a guard already built in above; they are named here so implementation can watch for them.

“FIRE IS DIFFERENT.”

The objection: wildland fire never built post-qualification evaluation, and the reasons are real. People fill these roles part-time, on top of a home job. Authority is deliberately spread across many positions. Outcomes are tangled up with weather, fuels, and terrain, so a good or bad outcome does not prove much about any one person. And there is no market keeping score. All true — and none of it prevents a solution. Aviation, nuclear power, and the military share most of these features and built evaluation systems anyway, by scoring how the person performs against a standard instead of scoring the outcome. Every gate in this framework does the same: behavior against a standard, never the incident's result (*Sustaining Decision Performance*).

PAPERWORK DRIFT

The most predictable failure. Any evaluated behavior slowly migrates toward its paperwork. Guard: thinking skills are scored only as live behavior. There is no form for the member to fill out — only score sheets for evaluators.

RELABELING PRESSURE

Staffing shortages will create pressure to advance people administratively — the exact mechanism that separated labels from capability at the team level before. Guard: the performance-only rule, written into doctrine, with any exception requiring a named decision and a scheduled review.

LOAD, NOT SKILL

The thinking skills are the first to fail under overload. Training them without managing workload installs a skill in conditions where it cannot run. Guard: the framework deploys together with workload review — span-of-control discipline, protected decision windows — and the lookout is treated as the unit's structural second set of eyes, not just a sensor.

NOT ENOUGH HARD ASSIGNMENTS

Mild seasons and quiet regions will starve the upper gates of hard-enough evaluated assignments, creating pressure to lower the bar rather than wait. Guard: scored simulation partly substitutes, and the exposure log records the gap honestly. The framework tolerates slower progression far better than it tolerates hollow certification.

EVALUATORS WHO CANNOT SEE THE GAPS

Behavior lists describe a standard; they do not give an evaluator the judgment to apply it. An evaluator pool short on hard-incident experience mirrors the member-side problem: evaluations that confirm what the member already believes, because both are calibrated to the same limited experience. Guard: the evaluator standard in 7.3, plus regular returns to operational roles for evaluators — evaluation skill fades the same way operational skill does. This risk sharpens as the appendices move up the organization: the pool of people who can credibly evaluate command and general staff positions is thin, and each appendix must say how it will handle that.

ADDING UP INDIVIDUALS IS NOT A TEAM

A CIMT is a team. Five Level 4 individuals do not automatically make a high-performing team, because team performance lives largely in the seams — between positions, between sections, and between the team and the organizations around it. This framework measures individuals. Reading a roster of individual levels as a statement of team capability repeats the exact configuration-equals-capability error the system has made before. Team capability needs its own measurement, and that is separate work.

9 · Rollout

Phase 1 — Build. Run the Section 5 method for the Division/Group Supervisor position (Appendix A). This is desk work plus an expert panel; no field burden.

Phase 2 — Pilot. Score the DIVS behavior lists against real candidate performance in existing simulation courses, and against a volunteer group of field evaluations across one

season, using the Appendix F instruments; agreement between evaluators is measured under the Appendix G protocol. Two questions: do the behaviors tell people apart, and do different evaluators agree? Revise.

Phase 3 — Deploy as development only. Publish the framework as a tool first. Members find their own level, receive focus areas, and build exposure plans — before any gate carries stakes. This builds the vocabulary and the habit while the cost of error is low.

Phase 4 — Attach stakes. Once the tool has a measured track record, gates begin governing assignment access at the upper levels, and evaluator and instructor roles at Level 5. Requirements follow measurement.

Phase 5 — Extend one appendix at a time. Only after the DIVS pilot proves the instrument does the next functional-area appendix carry any weight — gated on the same pilot logic. Appendices B through E are drafted in this document as working material for the expert panels; drafting them early lets the panels start while the pilot runs, but the pilot gate still governs when any of them becomes a requirement. Building five untested matrices at once multiplies risk instead of learning — which is exactly why the drafts stay drafts until the instrument is proven.

Command. The Incident Commander does not appear in this framework, and the absence is a sequencing decision, not an oversight. Command enters by the ladder: the Operations appendix extends from DIVS through Operations Branch Director to Operations Section Chief after the pilot proves the instrument, and Command sits above that extension. Until then, the certification rules in The Incident Tier govern the position, and the system-level case for evaluating command decision performance is carried in Sustaining Decision Performance. Building the Command appendix before the instrument is proven would put the framework's least-tested tool on its most consequential position.

End state. A CIMT workforce in which every member, in every section, can name their level, their focus areas, and the evidence their next gate needs — and a system that can answer, for the first time, where its bench actually sits on the path from qualified to expert.

10 · Corpus Anchors

Every idea this framework uses has a home in the corpus. The table locates them, so the framework can be checked against the doctrine it draws from — and so a revision in either place moves deliberately instead of by drift.

Framework element	Corpus source
Five-level structure; how the levels differ; where an experienced person re-enters	OLFG Ch. 10; Dreyfus and Dreyfus (1986)
Behavior-based rating scales; three-point standard; capability vs. compliance	OLFG Ch. 10; Smith and Kendall (1963)
Evaluator competence, calibration, and protection from grading pressure	OLFG Ch. 10

Framework element	Corpus source
Quick test; crystal ball drill and conflict variant; assumption audit; conflict count; changing the read	The Second Skill; Cohen, Freeman, and Thompson (1995); Klein (1993, 1998)
Developmental progression (slow, at speed, under stress, consistently); performance-only certification; exposure log; skill decay from underuse	The Incident Tier (DP-2, T-2, T-3, T-8)
System-level case for continued proficiency evaluation; cross-field precedent; the “fire is different” objection and its answer	Sustaining Decision Performance
Complete communication (SAPPER); intent two levels up; delegation to task	OLFG Chs. 1, 5
Double-meaning checks; softened-warning correction	OLFG Ch. 6
Trigger points; backup planning; executability check	OLFG Ch. 7
Structured reflection format (DEBrIEF); evidence that debriefs work	OLFG Ch. 8; Tannenbaum and Cerasoli (2013)
Noticing the new; stating reasons before acting; holding more than one read	OLFG Part I / Ch. 2
Three-tense situation picture plus warning signs; cognitive readiness safeguards	OLFG Ch. 11 (Endsley)
Simulation measurement infrastructure; behavior-marker validation	Teams and measurement literature review (IMT simulation testbeds; wargaming team measures)
Team capability is separate from summed individual capability	Teams and measurement literature review (multiteam-system seams); The Incident Tier
Trainee development and expertise-transfer methods	Developing-leaders literature review (PME best practices; expertise elicitation)

Appendix Template — What Every Appendix Must Contain

Each functional-area appendix is built with the Section 5 method and contains, in this order:

- The positions covered, and the ladder between them within the section — because Level 4–5 behavior in one position is the entry material for the next position up.
- The skill-area list, adjusted for the section — with the shared areas (Section 6) inherited, not rewritten.
- Level definitions in the section’s own operational language: what a Level 2 in this job actually looks like on shift.
- Focus areas per level: four or five, chosen by the cost of leaving the gap open.
- Gate evidence for each level, including the venues that supply it — what counts as a hard-enough assignment for this job, and what the simulation looks like.
- The evaluator pool: who can credibly evaluate this position, and what to do where the pool is thin.
- The empty cells, named honestly — the skill areas that genuinely start above Level 1 for this position.

Appendix A is the worked example. Match its structure.

Appendix A — Operations Section: Division/Group Supervisor

The Division/Group Supervisor (DIVS) appendix is built first, for two reasons. It is the position where the gap between qualification and expertise carries the most direct field consequence, and it is the position with the deepest evidence base in the corpus. It serves as the worked example for every appendix that follows.

Within Operations, DIVS is also the base of a ladder: DIVS to Operations Branch Director to Operations Section Chief. A future revision of this appendix will extend the backward mapping up that ladder, so that Level 4–5 behavior at DIVS reads as the entry material for the branch level.

Where a new DIVS starts. A newly qualified DIVS carries deep fire knowledge — fuels, weather, terrain, fire behavior, crew work — and that knowledge carries over. What does not carry over is the DIVS decision territory itself: supervising a division-scale operation, integrating multiple resources against an operational period plan, and carrying risk-acceptance authority at division scale. In that territory, the member starts at Level 1 and moves through the early levels faster than a true beginner would — but moves through them.

A.1 Level 1 — Newly Qualified: Working Through Structure

How the member operates

Runs the division by the book. Briefings follow standard formats, LCES is applied as written, span-of-control norms are followed as rules, and the operational period plan is executed as given. The member sees the cues but does not yet weight them well, and judgment calls are limited. The member performs reliably when the situation matches the plan, and calls for help when it does not. Calling early is the right behavior at this level — not a weakness.

Focus areas

- Complete communication. Every message that matters carries the situation, the action, the purpose, the plan, the execution, and the result — the SAPPER structure. Purpose is the piece that matters most, because it is the piece most often left out, and it is what lets the people below you adapt.
- Structured situation awareness. Build the picture in all three tenses — what was, what is, what is coming — and attach the fourth piece: what would tell you you're wrong. At this level, the warning signs mostly come from the plan and the supervisor. The skill being built is the habit of asking for them.
- Faithful execution with visible reasoning. Run the plan you were given, and be able to say back the intent behind it — from two levels above you — in your own words.

- Saying assumptions out loud. State what today's plan depends on, out loud, at the briefing and at trigger points. Smoothness is not expected yet; the habit is. The training tool is the crystal ball drill (Section 4), run slowly and out loud at this level.
- Asking for help at the right time. Recognize the edge of the current plan and say so early — without treating the call as a failure.

Method and gate

Method: do it slow. Moderate divisions, stable or predictable fire behavior, close contact with the OSC or branch, and a deliberate review before and after each shift. Gate to Level 2: a field evaluation on at least one full assignment of moderate difficulty, rated “meets” on every Level 1 behavior, with the focus-area behaviors confirmed by the evaluator. Time in grade proves nothing by itself.

A.2 Level 2 — Developing: Learning What Matters

How the member operates

Starts to weight cues. Recognizes recurring situation types — the division that is quiet because the work is done, versus quiet because something is about to change; the resource request that is routine, versus the one that signals a plan running past its assumptions. Still treats too many things as equally important, and is still at risk of plan lock: running a plan whose conditions have moved. The most common Level 2 failure is not missing the cue. It is seeing the cue and buying an explanation that keeps the current story alive.

Focus areas

- Building triggers and defending the math. Build trigger points that actually give the warning time they were drawn to give — and be able to defend the arithmetic: spread rate, travel time, margin. Do not inherit triggers as lines on a map.
- Naming the trade. Say what each tactical choice buys and what it costs, in words both a crew boss and a chief can act on.
- Noticing when the job outgrows the plan. Catch the moment the assignment is passing the plan's assumptions — before it becomes undeniable — and start the adjustment conversation instead of waiting for it.
- Purpose-driven resources. Order and assign resources against a stated purpose and priority — not out of habit, and not just because something is available.
- The quick test. Learn the rule for when to stop and double-check: high stakes, time available, unfamiliar situation. Apply it at set points — en route, at operational transitions, on arrival at trigger points.

Method and gate

Method: do it at speed. Full-tempo divisions at moderate difficulty; simulation and tactical decision games begin as a scored space for the thinking skills. Gate to Level 3: field evaluation at “meets” on the Level 2 behaviors, plus simulation scored on what can

be seen: assumptions said out loud per scenario, and conflicting cues acknowledged versus explained away.

A.3 Level 3 — Competent: Deliberate Division Management

How the member operates

Sees the division as a system with a multi-day path, not a series of shifts. Handles complexity through deliberate analysis and conscious planning: backup plans are built, not improvised; coordination with neighboring divisions and air operations is managed, not reacted to; decisions connect visibly to incident objectives. The work is effective and it takes effort — that effort is the mark of this level. Level 3 is the floor for independent assignment to complex divisions, and the natural center of most careers.

Focus areas

- Reasons stated before acting. Every big tactical choice carries its logic, said out loud before execution: if we do X, expect Y, because Z. That makes the reasoning checkable — and if Y does not happen, that becomes information instead of surprise.
- Backup plans built in advance. Alternatives identified, triggers named, resources positioned against the branches — during planning, not under pressure.
- The assumption audit and the conflict count. At set points, run the two-minute check: what does my current story assume, and how many conflicting cues have I explained away? Treat the second conflict as a signal, not a coincidence. The tools are the crystal ball drill and its conflict variant (Section 4), now run as trained habits.
- Communication under pressure. Fix softened language in both directions — hear the hedged warning from below for what it is, and strip the hedging from your own reports upward when the message matters.
- Managing the seams. Treat division boundaries, shared resources, and neighboring-division dependencies as first-order work, not administrative overhead.

Method and gate

Method: at speed, under stress. Assignments hard enough to feature real competition for resources, compressed decision windows, and consequential trade-offs — because skill shown only on easy divisions is assumed, not demonstrated. Gate to Level 4: field evaluation on a genuinely hard assignment, plus simulation with built-in surprises, scored on one behavior above all: when the injects contradict the member's story, does the member revise the story or defend it?

A.4 Level 4 — Proficient: Reading the Whole Division

How the member operates

Reads the division as a whole. Tells immediately what matters and what is noise, anticipates instead of reacting, and makes most decisions with little effort — which frees capacity for exactly the moments that demand it. The risk at this level mirrors the strength: fast pattern matching applied to a situation that has quietly stopped matching the pattern. The Level 4 focus areas exist mostly to audit that fluency.

Focus areas

- Noticing what is new. Catch the moment the situation is not what experience prepared you for — when the pattern match is running on surface looks — and deliberately shift from patterns to principles, slowing the planning tempo when newness exceeds confidence.
- Holding more than one read. Keep competing pictures of the situation open without settling too early. Decide as late as the situation allows, and structure resources and commitments to keep future options open.
- Dropping the story when it breaks. Show — not describe — the ability to let go of a working picture when the evidence turns against it, including pictures you built and briefed yourself.
- The OSC channel as an information system. Shape what moves up: believable resource requests, reports in priority order, and early warning that arrives while it is still cheap to act on.
- Running trainees deliberately. Structured exposure, reasoning made visible, practice corrected — not passengers.

Method and gate

Method: at speed, under stress, consistently. The gate shifts from one evaluated assignment to a record: several assignments at representative difficulty, each difficulty-coded, with performance held across all of them. Gate to Level 5: sustained “meets or exceeds” across the record, demonstrated story-changing behavior in both field and simulation, and evidence of trainee results.

A.5 Level 5 — Expert: Trusted Judgment, Checked

How the member operates

Runs on deep pattern recognition, saving deliberate analysis for genuine novelty and for the moments the quick test flags. The self-check runs as habit, and the expert runs it out loud, modeling the audit for others. Expertise at this level is an organizational asset — and it fades without exercise against truly hard conditions. Skills not run at their design load decay quietly while every routine assignment goes fine. The decay is documented at the

team level: organizations built by repeated exposure to the hardest incidents dull when their assignment diet drops below that level, and the fact that each individual assignment goes fine hides it (The Incident Tier, T-3). So Level 5 carries maintenance duties, not just standing.

Focus areas

- Developing others. The expert's main output shifts to the workforce: trainee progression, peer coaching, evaluator duty, and teaching from their own record — including the assignments that went sideways.
- Thinking out loud. The counter to a whole room confidently matching the same wrong pattern is the senior member who narrates their picture, their assumptions, and their warning signs in public — making the self-check a shared practice instead of a private one.
- Honest accounting of untested areas. Name where your expertise has not been tested by current conditions — new fuel regimes, new tempo, new coordination structures — and operate there the way a Level 3 would, not the way an expert would.
- Staying at design load. Seek, or be assigned, exposure at the difficulty your skills were built for. Where the season does not provide it, simulation and exercise design partly substitute — and the gap goes in the record instead of being assumed away.

Method and standing

Method: keep it sharp, pass it on, check yourself. Level 5 is a role change, not a finish line. Standing is maintained by a sustained performance record and by contribution to developing others — and it can be lost to decay. The evaluation rules of Section 7 apply to experts too.

A.6 Evaluator Pool and Empty Cells

The pool. DIVS is evaluated by qualified DIVS or above who carry a hard-incident record of their own, calibrated under the rules of Section 7.3. This is the deepest evaluator pool the framework will have; every appendix that follows works with less. The pool is maintained the same way member skill is: evaluators return to operational assignments on a schedule, because evaluation judgment decays the same way operational judgment does.

The empty cells. Two skill areas genuinely start above Level 1 for this position, and the matrix says so instead of inventing filler. Developing others begins at Level 3, where the member first carries enough spare capacity to run a trainee deliberately; before that, a trainee is a passenger. Formal evaluator duty begins at Level 5. Every other skill area, including risk-acceptance authority at division scale, is live from Level 1 — that authority is the reason this position anchors the framework.

Appendix B — Planning Section: Situation Unit Leader

Status. Draft for expert-panel review under Section 5, Step 4. Not yet piloted. Nothing in this appendix is a requirement until it passes the pilot logic of Section 9.

The Planning Section appendix anchors on the Situation Unit Leader (SITL), for the same two reasons Appendix A anchors on the DIVS: consequence and evidence. The situation picture is the one product every other part of the team consumes — the operations plan, the safety analysis, the cost projection all build on it. When the picture is wrong, everything built on it is wrong, and the people downstream usually cannot tell. The corpus evidence for situation assessment, warning signs, and picture-changing behavior maps onto this position more directly than onto any other in the section.

Within Planning, SITL is the base of a ladder: SITL to Planning Section Chief. The Resources Unit Leader is a parallel entry track into the same ladder; a future revision extends the backward mapping so that Level 4–5 behavior at the unit level reads as the entry material for the PSC.

Where a new SITL starts. A newly qualified SITL carries real fire knowledge — fire behavior, fuels, weather, mapping, and usually field time — and that carries over. What does not carry over is the SITL decision territory: running the situation function as a production system. Collecting against decision needs instead of collecting what is easy to get, validating what comes in, and saying plainly what the picture does not show. In that territory the member starts at Level 1 and moves through the early levels quickly — but moves through them.

B.1 Level 1 — Newly Qualified: Working Through Structure

How the member operates

Produces the standard products on the planning cycle: the incident map, the situation summary, the briefing displays. Field reports and remote sensing are taken mostly at face value. The member sees the information but does not yet weight the sources, and performs reliably when the inputs are good. When they are not, the member asks early — and at this level, flagging a hole in the picture is the job, not a failure.

Focus areas

- Complete products with a stated purpose. Every product names the decision it feeds. The SAPPER logic applies to information products the same way it applies to transmissions: purpose is the piece most often left out, and it is what lets the people using the product adapt when conditions move.
- The three-tense picture. Build the picture as what was, what is, and what is coming — and attach the fourth piece: what would say the picture is wrong. At this level the warning signs mostly come from the plan and the PSC; the skill being built is the habit of asking for them.

- Faithful production with visible reasoning. Build to the standard, and be able to say back which decisions each display supports — in your own words.
- Sourcing out loud. Label where each major element of the picture came from and how old it is, out loud, at the planning meeting. Smoothness is not expected yet; the habit is.
- Asking at the right time. When the data is stale or the reports conflict, say so early — before the picture goes into the briefing, not after.

Method and gate

Method: do it slow. Incidents with stable information flow, close contact with the PSC, and a deliberate review before and after each planning cycle. Gate to Level 2: a field evaluation across at least one full assignment of moderate complexity, rated “meets” on every Level 1 behavior, with the focus-area behaviors confirmed by the evaluator.

B.2 Level 2 — Developing: Learning What Matters

How the member operates

Starts to weight sources — the field observer reporting what they saw versus the one reporting what they infer; the IR pass that settles a question versus the one that only moves it. Still treats too many inputs as equal, and is still at risk of the planning-shop version of plan lock: a briefed picture that stays alive because updating it before the meeting is expensive. The most common Level 2 failure is not missing the conflicting report. It is seeing the report and buying an explanation that keeps the current map accurate.

Focus areas

- Confidence labels, defended. Attach a confidence level to each major element of the picture and be able to defend the basis: source, age, corroboration. Do not inherit confidence from the previous shift’s map.
- Naming the trade. Precision costs time; timeliness costs certainty. Say which one each product bought, in words both the OSC and the IC can act on.
- Noticing when the picture outruns its inputs. Catch the moment a display implies knowledge the data does not contain — before someone plans against it.
- Collection against decision needs. Task observers and request sensing against named decisions and trigger points — not out of routine, and not just because a platform is available.
- The quick test at production points. High stakes, time available, unfamiliar pattern — run the test before the picture goes to the briefing.

Method and gate

Method: do it at speed. Full planning tempo at moderate complexity; planning-cycle simulation begins as a scored space for the thinking skills. Gate to Level 3: field

evaluation at “meets” on the Level 2 behaviors, plus simulation scored on what can be seen: sourcing and assumptions said out loud per scenario, and conflicting inputs acknowledged versus explained away.

B.3 Level 3 — Competent: Deliberate Picture Management

How the member operates

Runs the situation function as a system across operational periods. Collection is planned ahead of need, validation loops are built in instead of improvised, and coordination with field observers, infrared, predictive services, and the GIS shop is managed rather than reacted to. The work is effective and it takes real effort — the defining mark of the level.

Focus areas

- Stating the projection’s reasoning before briefing it. The fire-behavior read carries its assumptions into the room: what it depends on, what it buys, what would change it.
- Backup collection with named triggers. When the IR platform is down or the observers are pulled, the fallback is already planned and the trigger that activates it is already named.
- Auditing your own picture at set points. Run the assumption audit on the current map at planned points in the cycle — not when a bad report forces it — and keep the conflict count.
- Briefing under pressure, both directions. Deliver the uncomfortable picture without softening it, and un-soften the field reports that arrive already hedged.
- Managing the seams. The boundary with Operations on trigger points, and with Safety on the hazard picture, is managed deliberately — not left to the planning meeting to sort out.

Method and gate

Method: at speed, under stress. Incidents with competing information streams and rapid transitions — complex enough to genuinely test the skills. Gate to Level 4: field evaluation on a genuinely hard assignment, plus simulation with built-in surprises, scored on one behavior above all: when the incoming data turns against the briefed picture, does the member change the picture or defend it?

B.4 Level 4 — Proficient: Reading the Whole Information System

How the member operates

Reads the incident’s information flow as a whole — where the picture is strong, where it is running on assumption, and where a quiet divergence between the map and the fire is opening. Anticipates information needs an operational period ahead. The risk mirrors the

strength: a confident, well-produced picture pattern-matched to an incident that has stopped matching it.

Focus areas

- Noticing what is new. New fuel behavior, new tempo, new spread pattern — catch when the projection is running on surface resemblance, and slow the production tempo when newness exceeds confidence.
- Holding more than one picture. Keep competing reads of fire potential open in the products themselves — ranges and scenarios, not a single confident line.
- Dropping the picture when it breaks. Show — not describe — the ability to retract a briefed projection when the evidence turns against it, including one you built and defended in the meeting.
- The PSC and IC channel as an information system. Shape what moves up: prioritized, believable, and early enough that acting on it is still cheap.
- Running SITL trainees deliberately. Structured exposure, reasoning made visible, practice corrected — not passengers.

Method and gate

Method: at speed, under stress, consistently. Gate to Level 5: a record across several complex incidents, each difficulty-coded, with sustained “meets or exceeds,” demonstrated picture-changing behavior in both field and simulation, and evidence of trainee results.

B.5 Level 5 — Expert: Trusted Picture, Checked

How the member operates

Runs on deep pattern recognition of fire potential and information quality, saving deliberate analysis for genuine novelty and for the moments the quick test flags. Runs the picture audit out loud in the planning meeting, modeling the habit for the room. The decay logic of Appendix A applies unchanged: a run of ordinary incidents dulls exactly the projection skills the hard ones require, and every fine outcome hides it.

Focus areas

- Developing others. Trainee progression, peer coaching, evaluator duty, and teaching from your own record — including the projections that were wrong.
- Thinking out loud. Narrate the picture, its sourcing, its assumptions, and its warning signs in public — the counter to a whole planning meeting confidently agreeing on the same wrong map.
- Honest accounting of untested areas. New fuel regimes, new sensing tools, new coordination structures — name where the expertise has not been tested, and operate there the way a Level 3 would.

- Staying at design load. Seek exposure at the complexity the skills were built for; where the season does not provide it, simulation partly substitutes, and the gap goes in the record.

Method and standing

Method: keep it sharp, pass it on, check yourself. Standing is maintained by a sustained record and by contribution to developing others, and it can be lost to decay. Section 7 applies to experts too.

B.6 Evaluator Pool and Empty Cells

The pool. Qualified SITLs and PSCs with complex-incident records — thinner than the DIVS pool. Handling: cross-team calibration on recorded planning-cycle scenarios, and any evaluation done by someone who has not held the position is marked as such in the record, per Section 7.3.

The empty cells. Developing others begins at Level 3, and formal evaluator duty at Level 5, as in Appendix A. Risk-acceptance authority at life-safety scale is not part of this position's territory; the risk this position carries is the honesty of the picture, and the behavior lists score it there. The matrix says so instead of inventing filler.

Appendix C — Logistics Section: Logistics Section Chief

Status. Draft for expert-panel review under Section 5, Step 4. Not yet piloted. Nothing in this appendix is a requirement until it passes the pilot logic of Section 9.

The Logistics appendix anchors at the section chief, because the decision territory that separates qualified from expert in this section lives at the chief's desk: forecasting demand ahead of the incident instead of reacting to it, and making long-lead-time commitments against a future that has not arrived yet. The unit leaders below — Supply, Facilities, Ground Support, Communications, Food, Medical — form the ladder into the position, through the Service and Support Branch Directors. A future revision extends the backward mapping down that ladder, so that Level 4–5 behavior at the unit level reads as LSC entry material.

Where a new LSC starts. Deep procedural knowledge — ordering systems, agreements, unit operations — carries over. What does not carry over is the forecasting territory: reading the incident's trajectory as a demand curve, and committing resources against it before the demand is undeniable. In that territory the member starts at Level 1.

C.1 Level 1 — Newly Qualified: Working Through Structure

How the member operates

Runs the section by the structure. Orders are filled as they arrive, status is tracked by the book, and the support plan is executed as given. The member sees the demand signal but does not yet weight it, and escalates early when demand outruns capacity — the right behavior at this level.

Focus areas

- Complete communication. Status reports and orders that carry situation, action, and purpose — because purpose is what lets ordering managers and unit leaders adapt when the exact item or timeline is not available.
- The three-tense logistics picture. What was ordered, what is on hand, what is inbound — and the fourth piece: what would say the forecast is wrong. At this level the warning signs come from the support plan and the incident's ordering history; the habit being built is asking for them.
- Faithful execution with visible reasoning. Run the support plan as given, and be able to say back what the operation is trying to do — so the section supports intent, not just orders.
- Assumptions out loud. Lead times, vendor capacity, transport windows — said out loud at the planning meeting, not discovered at the shortage.
- Asking at the right time. When the demand signal outruns capacity, say so early — the call is the job, not a failure.

Method and gate

Method: do it slow. Assignments of moderate complexity with close IC or deputy contact and a deliberate review each cycle. Gate to Level 2: field evaluation on at least one full assignment, “meets” on every Level 1 behavior.

C.2 Level 2 — Developing: Learning What Matters

How the member operates

Starts to weight orders — the routine request versus the one that signals the operation is shifting. Still at risk of the logistics form of plan lock: servicing the written support plan while the incident moves past it. The most common Level 2 failure is not missing the surge signal. It is seeing the signal and accepting the explanation that keeps the current ordering posture adequate.

Focus areas

- Lead-time triggers and defending the math. Set ordering points that actually give the lead time they were set to give — and defend the arithmetic: consumption rate, transport time, margin.
- Naming the trade. What pre-positioning costs versus what a stockout costs, in words both Finance and Operations can act on.
- Noticing when demand outgrows the plan. Catch the moment the incident is passing the support plan’s assumptions — before the shortage makes it undeniable — and start the adjustment conversation.
- Ordering against purpose. Commit resources against a stated purpose and priority — not habit, and not just because a source is available.
- The quick test at commitment points. Run the test before the long-lead order goes out, not after.

Method and gate

Method: do it at speed. Full-tempo assignments; simulation begins as a scored space — see C.6 on the simulation venue. Gate to Level 3: field evaluation plus scored simulation: assumptions said out loud, and surge signals acknowledged versus explained away.

C.3 Level 3 — Competent: Deliberate Support Management

How the member operates

Runs logistics as a system across operational periods. Demand is projected ahead, contingency supply is built before it is needed, and the seams — Operations for the demand signal, Planning for the projection, Finance for the commitment — are managed rather than reacted to. Effective, with real effort: the mark of the level.

Focus areas

- Stating the demand reasoning before the meeting. The forecast carries its assumptions into the room: what it depends on, what it buys, what would change it.
- Contingency support with named triggers. The backup source, route, or facility is planned before it is needed, with the trigger that activates it already named.
- Auditing the forecast at set points. Run the assumption audit on the demand projection at planned points in the cycle, and keep the conflict count on surge signals explained away.
- Delivering constraints without softening. The “we cannot support that” call, made early and plainly — and the hedged unit-leader report heard for what it is.
- Managing the seams. The boundaries with Operations, Planning, and Finance are worked deliberately, because the forecast is only as good as the demand signal crossing them.

Method and gate

Method: at speed, under stress. Assignments complex enough to genuinely test the forecasting — rapid growth, competing incidents, constrained supply. Gate to Level 4: hard-assignment field evaluation plus simulation with surprises, scored above all on: when the numbers turn against the forecast, does the member change the forecast or defend it?

C.4 Level 4 — Proficient: Reading the Demand Curve***How the member operates***

Reads the incident’s logistics trajectory as a whole — anticipates the demand curve instead of tracking it, and makes most commitment calls with little effort, freeing capacity for the exceptions. The risk mirrors the strength: a demand pattern matched to an incident that has quietly changed shape.

Focus areas

- Noticing what is new. A fuel regime, tempo, or jurisdictional mix the old consumption models do not fit — catch it, and slow the commitment tempo when newness exceeds confidence.
- Holding more than one demand read. Keep competing demand scenarios open and structure commitments to keep future options open — flexible contracts over fixed ones when the read is uncertain.
- Dropping the forecast when it breaks. Show the ability to retract a projection when the actuals turn against it — including one already briefed and budgeted.
- The command channel as an information system. Believable projections, constraint warnings that arrive while alternatives still exist, and reports in priority order.

- Running trainees deliberately. Structured exposure, reasoning made visible, practice corrected.

Method and gate

Method: at speed, under stress, consistently. Gate to Level 5: a record across several complex assignments, sustained “meets or exceeds,” demonstrated forecast-changing behavior, and evidence of trainee results.

C.5 Level 5 – Expert: Trusted Forecast, Checked

How the member operates

Runs on deep pattern recognition of incident demand, with the self-check running as habit and out loud – the forecast audit narrated in the planning meeting, so others learn it by watching. The decay logic applies: forecasting skill built on complex incidents dulls on a diet of ordinary ones, and every adequately supported assignment hides it.

Focus areas

Developing others as the primary output; narrating the forecast audit in public; honest accounting of untested areas – new contract structures, new supply environments – and operating there the way a Level 3 would; staying at design load, with simulation substituting where the season does not provide and the gap recorded.

Method and standing

Method: keep it sharp, pass it on, check yourself. Standing is maintained by record and contribution, and can be lost to decay. Section 7 applies to experts too.

C.6 Evaluator Pool and Empty Cells

The pool. Very thin – few people hold this position at complex-incident level in any single geographic area. Handling: cross-geographic-area calibration on recorded scenarios, and non-holder evaluations marked in the record, per Section 7.3.

The empty cells. Life-safety risk-acceptance authority is largely absent from this position’s territory; the risk carried here is operational continuity and fiscal commitment, and the behavior lists score it there. Developing others begins at Level 3. One honest gap in the venues: the existing simulation infrastructure has no logistics track – the planning-cycle and sand-table exercises test Operations and Planning, not the demand forecast. Building that track is a named cost of this appendix, not something assumed to exist.

Appendix D — Finance/Administration Section: Finance/Admin Section Chief

Status. Draft for expert-panel review under Section 5, Step 4. Not yet piloted. Nothing in this appendix is a requirement until it passes the pilot logic of Section 9.

The Finance/Administration appendix anchors at the section chief. The gap that matters in this section is the difference between cost as bookkeeping and cost as decision support. A qualified FSC can account for the incident. An expert FSC can tell the IC and the agency administrator what a strategy will cost before it is chosen, and what exposure is accumulating while it runs. The unit leaders — Time, Procurement, Compensation/Claims, Cost — form the ladder into the position; a future revision extends the backward mapping down it.

Where a new FSC starts. Procedures, agreements, and systems carry over. What does not is the projection territory: turning transaction streams into a forward-looking picture the command team can act on. In that territory the member starts at Level 1.

D.1 Level 1 — Newly Qualified: Working Through Structure

How the member operates

Processes by procedure — accurate and on schedule. Cost drivers are taken as given, and the member escalates early when exposure appears. Escalating early is the right behavior at this level.

Focus areas

- Complete reporting with a stated purpose. Every report names the decision it feeds — a cost summary built for the agency administrator's strategy call is a different product than one built for the ledger.
- The three-tense fiscal picture. What is obligated, what is on hand, what is projected — and the fourth piece: what would say the projection is wrong.
- Faithful execution with visible reasoning. Run the procedures as written, and be able to say back what the incident is trying to do — so the section informs intent, not just compliance.
- Assumptions out loud. Rate assumptions, cost-share terms, duration assumptions — said out loud where the command team can hear them.
- Asking at the right time. When an exposure appears — a claims pattern, a procurement gap — say so early.

Method and gate

Method: do it slow. Moderate assignments, close contact with the IC or deputy, deliberate review each cycle. Gate to Level 2: field evaluation, "meets" on every Level 1 behavior.

D.2 Level 2 — Developing: Learning What Matters

How the member operates

Starts to weight transactions — the routine entry versus the one that carries a signal: the claims pattern that flags an operational problem, the procurement request that signals the plan drifting past its assumptions. The Level 2 risk is accurate books on a diverging incident: every transaction correct, and the picture they add up to unexamined.

Focus areas

- Projection triggers and defending the math. Set cost-projection review points that actually give warning time — and defend the arithmetic: burn rate, commitment schedule, margin.
- Naming the trade. What speed of support costs in exposure, and what tight control costs in operational tempo — said in words the IC can act on.
- Noticing when the incident outgrows the cost picture. Catch the moment the projection's assumptions have moved — before the variance makes it undeniable.
- Procurement against purpose. Commit against stated purpose and priority, not routine.
- The quick test before large commitments. High stakes, time available, unfamiliar structure — run the test before signing, not after.

Method and gate

Method: do it at speed. Full-tempo assignments; scored simulation per D.6. Gate to Level 3: field evaluation plus scored simulation: assumptions said out loud, and signal transactions acknowledged versus explained away.

D.3 Level 3 — Competent: Deliberate Fiscal Management

How the member operates

Runs the section as a system across the incident's life. Projection is deliberate, cost-share and claims exposure are managed ahead of need, and the seams — Logistics for commitments, Command for strategy cost, the agency administrator for delegation limits — are worked rather than reacted to.

Focus areas

- Stating the projection's reasoning before briefing it. The cost picture carries its assumptions into the room.
- Contingency fiscal plans with named triggers. The cost-share renegotiation, the added jurisdiction, the demobilization surge — planned before they arrive, with triggers named.

- Auditing assumptions at set points. Run the audit on the projection at planned points, and count the variances explained away.
- Delivering unwelcome numbers without softening. The projection the strategy does not want to hear, delivered early and plainly.
- Managing the seams. Especially the agency-administrator seam, where the cost picture meets the strategic decision.

Method and gate

Method: at speed, under stress. Assignments with cost-share, multiple jurisdictions, and active claims load. Gate to Level 4: hard-assignment field evaluation plus simulation with surprises, scored on: when the actuals turn against the projection, does the member change the projection or defend it?

D.4 Level 4 — Proficient: Reading the Fiscal Trajectory

How the member operates

Reads the incident's fiscal trajectory as a whole and anticipates it — exposure named before it accumulates, projection revised before the variance forces it. The risk mirrors the strength: a projection pattern-matched to an incident whose cost structure has quietly changed.

Focus areas

- Noticing what is new. Contract types, jurisdictional mixes, and cost structures the old models do not fit.
- Holding more than one projection. Ranges and scenarios in the products, not a single confident number.
- Dropping the projection when it breaks. Including one already briefed to the agency administrator.
- The command channel as an information system. Exposure warnings that arrive while options still exist, in priority order, in believable form.
- Running trainees deliberately.

Method and gate

Method: at speed, under stress, consistently. Gate to Level 5: a record across several complex assignments, sustained “meets or exceeds,” demonstrated projection-changing behavior, and trainee evidence.

D.5 Level 5 — Expert: Trusted Numbers, Checked

How the member operates

Deep pattern recognition on incident cost behavior, with the audit run as habit and narrated in public. Decay logic applies: projection skill built on complex cost-share incidents dulls on a diet of simple ones.

Focus areas

Developing others as the primary output; narrating the projection audit in public; honest accounting of untested areas — new agreement structures, new cost environments — and operating there as a Level 3 would; staying at design load, with the gap recorded where the season does not provide it.

Method and standing

Method: keep it sharp, pass it on, check yourself. Section 7 applies to experts too.

D.6 Evaluator Pool and Empty Cells

The pool. Thin, as with Logistics. Handling: cross-team calibration on recorded scenarios; non-holder evaluations marked, per Section 7.3.

The empty cells. Named honestly, this position's thinking skills run at a lower tempo than the field positions', so "at speed, under stress" is tested in compressed planning cycles and simulation rather than tactical tempo — and the behavior lists are written for that tempo, not borrowed from Operations. A hard-enough assignment for this position means cost-share, multiple jurisdictions, and active claims load, and the exposure log codes for those features. Life-safety risk territory is absent; exposure territory is scored instead. Developing others begins at Level 3. The simulation venue must be built, as in Appendix C — a named cost.

Appendix E — Command Staff: Safety Officer

Status. Draft for expert-panel review under Section 5, Step 4. Not yet piloted. Nothing in this appendix is a requirement until it passes the pilot logic of Section 9.

The Command Staff appendix anchors on the Safety Officer, the closest command-staff analog to the DIVS decision territory: a field-facing risk read, and the authority to alter or stop an operation on the strength of it. The Public Information Officer and Liaison Officer share the level structure, but their evidence base in the corpus is thin; their behavior lists are extensions to be built panel-first, with the heavier reliance on Section 5, Step 4 that Section 8 requires for exactly this situation.

Where a new Safety Officer starts. Hazard knowledge, safety systems, and field experience carry over. What does not is the incident-scale risk territory: reading risk across the whole operation rather than one piece of line, and carrying the stop-work call as a judgment rather than a rule. In that territory the member starts at Level 1.

E.1 Level 1 — Newly Qualified: Working Through Structure

How the member operates

Applies the safety systems as written — LCES review, standard mitigations, checklist audits. Sees hazards but weights them evenly, and judgment calls are limited. Escalates early when the situation moves past the safety plan — the right behavior at this level.

Focus areas

- Complete communication. A hazard report that carries situation, action, and purpose — because purpose is what lets the DIVS act on it instead of filing it.
- The three-tense risk picture. What was, what is, what is coming — plus what would say the risk read is wrong. At this level the warning signs come from the IAP and the OSC; the habit being built is asking for them.
- Faithful application with visible reasoning. Apply the mitigations as written, and be able to say back why each one exists — in your own words.
- Assumptions out loud. What today's safety plan depends on — said at the briefing and at trigger points.
- Asking at the right time. Recognize the edge of the safety plan and say so early.

Method and gate

Method: do it slow. Moderate assignments, stable conditions, close OSC contact, deliberate review each shift. Gate to Level 2: field evaluation, “meets” on every Level 1 behavior.

E.2 Level 2 — Developing: Learning What Matters

How the member operates

Starts to weight hazards and recognize situation types — the division quiet because the work is done, versus quiet before a change. This seat carries a risk the others do not carry as sharply: softening the warning to preserve the working relationship. The most common Level 2 failure here is not missing the hazard. It is seeing it and delivering it in words easy to discount.

Focus areas

- Trigger math. Verify that the trigger actually buys the escape time it was drawn to buy — spread rate, travel time, margin — rather than inheriting it as a line on the map.
- Naming the trade. What the mitigation costs the operation, said plainly — because a risk decision made on a hidden cost is not an honest one.
- Noticing when operations outgrow the safety plan. Catch the moment before it becomes undeniable, and start the conversation.
- Mitigation against purpose. Recommend mitigations against the stated risk, not reflex — the mitigation that does not serve a purpose costs credibility that the important call will need later.
- The quick test. En route, at operational transitions, at trigger points.

Method and gate

Method: do it at speed. Full-tempo assignments; simulation begins as a scored space. Gate to Level 3: field evaluation plus scored simulation — with un-softened delivery of warnings scored explicitly, alongside assumptions said out loud and cues acknowledged versus explained away.

E.3 Level 3 — Competent: Deliberate Risk Management

How the member operates

Runs the safety function across the incident as a system. Hazard analysis is deliberate, presence is planned against where risk concentrates rather than where travel is convenient, and the seams with the DIVS, the OSC, and Planning are managed rather than reacted to.

Focus areas

- Stating the risk reasoning before the period starts. The risk read carries its assumptions into the briefing: what it depends on, what would change it.

- Contingency risk plans with named triggers. The escalation, the weather change, the added exposure — planned ahead, with the trigger and the response already named.
- Auditing your own risk read at set points. Planned audit points, and the conflict count on hazards explained away.
- Un-softening in both directions. Hear the hedged field report for what it is, and send the uncomfortable call plainly — including the stop-work call. This is the load-bearing skill of the position, and the gate scores it directly.
- Managing the seams. The DIVS relationship especially: influence that works through credibility and clarity, because the position mostly does not command.

Method and gate

Method: at speed, under stress. Assignments hard enough to genuinely test the read. Gate to Level 4: hard-assignment field evaluation plus simulation with surprises, scored on: when the facts turn against the risk read, does the member change the read or defend it — and does the warning go out un-softened?

E.4 Level 4 — Proficient: Reading the Incident’s Risk Trajectory

How the member operates

Reads the incident’s whole risk trajectory — where tomorrow’s exposure concentrates, which divisions are running on margin, where the operation’s tempo is outrunning its safeguards. Anticipates instead of reacting. The risk of the level is fluent pattern matching on an incident that has changed.

Focus areas

- Noticing what is new. Fuel regimes, tempo, and coordination structures the pattern library was not built on — and slowing down when newness exceeds confidence.
- Holding more than one risk read. Competing reads of where the exposure sits, held open without settling early.
- Dropping the read when it breaks. Including a risk picture already briefed to command.
- Shaping what moves up. Risk reporting in priority order, early warning that arrives while acting on it is cheap, and credibility spent where it matters.
- Running trainees deliberately.

Method and gate

Method: at speed, under stress, consistently. Gate to Level 5: a record across several hard assignments, sustained “meets or exceeds,” demonstrated read-changing behavior in field and simulation, and trainee evidence.

E.5 Level 5 — Expert: Trusted Judgment, Modeled

How the member operates

Runs on deep pattern recognition, with the self-check as habit and out loud. For this position the public narration carries extra weight: the stop-work judgment modeled in the open — the reasoning, not just the call — is how the next generation learns that the call is a judgment, made on defensible grounds, and safe to make. Teaching includes the member's own record, including the calls that were wrong in both directions: the stop that was not needed, and the go that should have been a stop.

Focus areas

Developing others as the primary output; thinking out loud, with the risk audit narrated in public; honest accounting of untested areas and operating there as a Level 3 would; staying at design load, with simulation substituting where the season does not provide, and the gap recorded.

Method and standing

Method: keep it sharp, pass it on, check yourself. Section 7 applies to experts too.

E.6 Evaluator Pool and Empty Cells

The pool. Among the thinnest in the framework — few evaluators have operated as Safety Officer at complex-incident level. Handling: the Section 7.3 rules, cross-team calibration on recorded scenarios, and non-holder evaluations marked in the record. This is the case Section 8 names directly, and this appendix accepts the burden it assigns.

The empty cells. Resource and workload management is thin for this position and the list says so, rather than importing behaviors from Operations to fill the cell. Developing others begins at Level 3. Command authority is absent by design — the position works through credibility, and the behavior lists score influence, clarity, and the un-softened call instead. The PIO and Liaison Officer extensions inherit the level structure from this appendix; their behavior lists start with the Step 4 panel, not with a draft.

Appendix F — Pilot Instrument Set

This appendix contains the instruments Phases 1 and 2 run on. Section 7 describes the measurement system; this appendix builds it. Five instruments: the DIVS rating sheets (F.1), the simulation score sheet (F.2), the exposure log and difficulty coding scheme (F.3), the structured reflection template (F.4), and the gate record (F.5). Each consumes the others: the rating sheets and score sheet supply the evidence, the coding scheme says what the evidence was tested against, the reflection carries the member's own account, and the gate record brings the three streams together under the Section 7.3 rules.

One rule governs the whole set. These instruments measure; they do not yet require. Nothing scored here carries stakes until the pilot proves the instrument — the behaviors tell people apart, and different evaluators agree.

F.1 — DIVS Rating Sheets

One sheet per gate, Levels 1 through 4, plus the Level 5 standing review. Each sheet converts the Appendix A focus areas into behaviorally anchored items on the Section 7.1 three-point scale. The evaluator's question is never how the performance felt. It is which of these behaviors did I see.

How to score. Rate every item from observed behavior on the assignment or scenario, and note the evidence — what was seen, where, when — in one line per item. A member's current level is wherever they rate meets on every item that applies. The gaps to the next level become their focus areas on the gate record (F.5). The safeguards item is standing: it appears on every sheet and is scored as safeguards in use, never as self-report.

The anchors below are drafts for the Step 4 panel. The panel's test is unchanged: a behavior every qualified DIVS already shows, or one no evaluator could observe in the field, gets rewritten or cut.

Sheet 1 — Level 1 gate (advancement to Level 2)

Minimum exposure: one full assignment coded D1 or higher (see F.3). Method context: do it slow.

Behavior	Does not meet	Meets	Exceeds
Complete communication (SAPPER)	Messages that matter arrive missing purpose or result; the receiver has to ask.	Messages that matter carry situation, action, purpose, plan, execution, and result. Purpose stated without prompting.	SAPPER holds under time pressure; subordinates can say the purpose back.
Structured situation awareness	Picture is present-tense only; no warning signs named or asked for.	Builds the picture in three tenses and asks the plan or the supervisor for the warning signs.	Proposes warning signs of their own and checks them at trigger points.
Faithful execution with visible reasoning	Executes without being able to state the intent behind the plan — or departs from the plan without a call.	Runs the plan as given and states the intent from two levels up in their own words.	Flags where the plan and the intent could pull apart — before executing.
Saying assumptions out loud	No assumptions voiced at the briefing or at trigger points.	States what the plan depends on, out loud, at the briefing and at trigger points — slowly	Runs the crystal ball drill unprompted and names the risks it turned up.

Behavior	Does not meet	Meets	Exceeds
		is fine; the habit is the standard.	
Asking for help at the right time	Rides past the edge of the plan without a call — or calls constantly without reading the edge.	Names the edge of the current plan and calls early, treating the call as procedure, not failure.	The call arrives with the situation, the edge crossed, and options.
Safeguards in use (standing item, every gate)	No go/no-go use; critical calls made solo when worn down; big decisions at circadian low points without concurrence.	Uses go/no-go checklists; seeks a second person's concurrence on critical calls; keeps big decisions away from body-clock low points when practicable.	Builds safeguard use into the division's rhythm and prompts it in others.

Sheet 2 — Level 2 gate (advancement to Level 3)

Minimum exposure: assignments coded D2 or higher, plus scored simulation (F.2). Method context: do it at speed.

Behavior	Does not meet	Meets	Exceeds
Building triggers and defending the math	Inherits trigger points as lines on a map; cannot defend the arithmetic when asked.	Builds triggers that give the warning time they were drawn for, and defends the math: spread rate, travel time, margin.	Adjusts triggers in-shift as rates change and briefs the change.
Naming the trade	States choices without cost — what the choice buys and gives up goes unsaid.	Says what each tactical choice buys and what it costs, in words both a crew boss and a chief can act on.	Names the trade before being asked — including the option not taken.
Noticing when the job outgrows the plan	Explains cues away until the mismatch is undeniable.	Catches the assignment passing the plan's assumptions and starts the adjustment conversation early.	Names which assumption failed and arrives with the adjustment.
Purpose-driven resources	Orders by habit, or because something is available.	Orders and assigns resources against a stated purpose and priority.	Releases what the purpose no longer needs.
The quick test	No observable stop-and-check at transitions or trigger points.	Applies the rule — high stakes, time available, unfamiliar — at set points: en route, at transitions, on arrival at triggers.	Says out loud which branch of the rule applies, and why.
Safeguards in use (standing item, every gate)	No go/no-go use; critical calls made solo when worn down; big decisions at circadian low points without concurrence.	Uses go/no-go checklists; seeks a second person's concurrence on critical calls; keeps big decisions away from body-clock low points when practicable.	Builds safeguard use into the division's rhythm and prompts it in others.

Sheet 3 — Level 3 gate (advancement to Level 4)

Minimum exposure: a genuinely hard assignment — coded D3 — plus simulation with built-in surprises.

Method context: at speed, under stress.

Behavior	Does not meet	Meets	Exceeds
Reasons stated before acting	Big tactical choices executed with the logic unstated.	Voices the logic before execution on every big choice: if we do X, expect Y, because Z.	When Y does not happen, treats it as information and says so.
Backup plans built in advance	Alternatives improvised under pressure.	Alternatives identified, triggers named, resources positioned against the branches — during planning.	Checks each backup for executability — time and resources — before committing to it.
The assumption audit and the conflict count	Conflicting cues explained away with no running count.	Runs the two-minute audit at set points and treats the second conflict as a signal, not a coincidence.	Narrates the count for the division and acts on it.
Communication under pressure	Hedged warnings pass through hedged; own reports soften as stakes rise.	Hears the hedged warning from below for what it is; strips the hedging from upward reports that matter.	Corrects softened language in others, in real time, in both directions.
Managing the seams	Boundaries, shared resources, and neighbor dependencies treated as overhead.	Seams managed as first-order work: dependencies known, contacts made, conflicts surfaced.	Raises seam risk to the OSC before it binds.
Safeguards in use (standing item, every gate)	No go/no-go use; critical calls made solo when worn down; big decisions at circadian low points without concurrence.	Uses go/no-go checklists; seeks a second person's concurrence on critical calls; keeps big decisions away from body-clock low points when practicable.	Builds safeguard use into the division's rhythm and prompts it in others.

Sheet 4 — Level 4 gate (advancement to Level 5)

Minimum exposure: a record — several assignments at D3–D4, each coded, with performance held across all of them, plus trainee results. Method context: at speed, under stress, consistently.

Behavior	Does not meet	Meets	Exceeds
Noticing what is new	Applies the pattern on a surface match; planning tempo never slows.	Names the mismatch out loud, shifts from patterns to principles, and slows the tempo when newness exceeds confidence.	Makes the naming a room practice — invites challenge to their own read.
Holding more than one read	Settles on one story early and commits resources to it.	Keeps competing reads open, decides as late as the situation allows, and structures commitments to keep options open.	Names the information that would separate the reads — and when it arrives.

Behavior	Does not meet	Meets	Exceeds
Dropping the story when it breaks	Defends a briefed picture against the evidence — especially their own.	Demonstrably lets go of a working picture when the evidence turns, including pictures they built and briefed.	Re-briefs the new picture without defending the old one.
The OSC channel as an information system	Raw or unprioritized traffic upward; warning arrives late.	Believable resource requests, reports in priority order, early warning that arrives while it is still cheap to act on.	Shapes the channel so the OSC's picture needs no repair.
Running trainees deliberately	The trainee is a passenger.	Structured exposure, reasoning made visible, practice corrected.	The trainee's record shows progression attributable to the plan.
Safeguards in use (standing item, every gate)	No go/no-go use; critical calls made solo when worn down; big decisions at circadian low points without concurrence.	Uses go/no-go checklists; seeks a second person's concurrence on critical calls; keeps big decisions away from body-clock low points when practicable.	Builds safeguard use into the division's rhythm and prompts it in others.

Sheet 5 — Level 5 standing review

Run at the Section 7.4 schedule. Level 5 is standing, not tenure: it is maintained by this review and can be lost to decay.

Behavior	Does not meet	Meets	Exceeds
Developing others	Standing without output: no trainee progression, no evaluator duty, no teaching.	Main output visible in the workforce: trainee progression, peer coaching, evaluator duty, teaching from their own record — including the assignments that went sideways.	Builds development capacity in others — evaluators and coaches, not just operators.
Thinking out loud	The self-check runs privately, if at all; the room never hears the read.	Narrates the picture, the assumptions, and the warning signs in public — making the audit a shared practice.	The narration changes the room's behavior: others start voicing their own.
Honest accounting of untested areas	Expertise claimed uniformly; no untested territory named.	Names where expertise has not been tested by current conditions — new fuels, new tempo, new structures — and operates there as a Level 3 would.	The accounting appears in the record and shapes their own exposure plan.
Staying at design load	Assignment diet below design load, with no gap recorded.	Seeks or accepts D4 exposure; where the season does not provide it, simulation partly substitutes — and the gap goes in the record.	Designs the exercises that supply design load to others.

Behavior	Does not meet	Meets	Exceeds
Safeguards in use (standing item, every gate)	No go/no-go use; critical calls made solo when worn down; big decisions at circadian low points without concurrence.	Uses go/no-go checklists; seeks a second person's concurrence on critical calls; keeps big decisions away from body-clock low points when practicable.	Builds safeguard use into the division's rhythm and prompts it in others.

F.2 — Simulation Score Sheet

Simulation is where the thinking skills get measured, because field observation only catches them by luck. The scenario's difficulty code is set by design before the exercise runs (F.3, rule 3) — the designer decides the band and features, and the sheet records them. The behaviors scored are the Section 4 tools made visible.

Header

Field	Entry
Scenario ID	
Designed code	Band and features set at design, e.g. D3-TNI
Member / position	
Gate targeted	Level 2 · 3 · 4 · 5 standing
Evaluator	Name · holder / non-holder (marked per Section 7.3)

Scored behaviors

Behavior	What is counted	Record
Assumptions said out loud	Count per scenario. An assumption counts when it names something the plan depends on — not a restatement of the plan.	Count: ____ Notes:
Conflicting cues — acknowledged vs. explained away	Two counts. Acknowledged: the member names the cue as conflicting. Explained away: the cue is absorbed into the current story. Also record whether the member tracked their own count.	Ack: ____ Explained: ____ Tracked own count: Y / N
Story revision on the turn	When injects contradict the member's story: revised (new story built) · patched (old story amended) · defended. This is the single most heavily weighted item at the Level 3 and 4 gates.	Revised / Patched / Defended At which inject: ____
Reasons stated before acting	Instances of if-X-expect-Y-because-Z voiced before execution on major choices.	Count: ____

Behavior	What is counted	Record
Trigger arithmetic defended	When challenged, the member defends the math: spread rate, travel time, margin.	Y / N · not challenged
Quick test applied	At the designed decision points: did the member stop and check when stakes, time, and unfamiliarity called for it — and act on the pattern when they did not?	Applied: ____ of ____ designed points

Per-level standards

Numeric thresholds are set by the panel and validated in the pilot — requirements follow measurement. The qualitative standard per gate:

- **Level 2 gate.** Assumptions voiced at [panel] per scenario, even if slow; second conflict treated as a signal when prompted.
- **Level 3 gate.** Audit and count run unprompted at set points; story revised on the turn without prompting.
- **Level 4 gate.** The Level 3 standard held across a scenario set, plus narration — the member runs the audit out loud.
- **Level 5 standing.** Design-load scenarios only (D4). The standard is the Level 4 standard plus the member's narration changing evaluator-visible behavior in the room.

F.3 — Exposure Log and Difficulty Coding Scheme

Every evaluated assignment gets one code. The code records what the assignment actually demanded of the member at their position — not what the incident was called. A quiet division on a Type 1 incident codes as a quiet division. The code exists so the record is honest about what the member has been tested against, and so no gate passes on exposure that never stressed the behaviors being certified.

The code has two parts: a band and features.

The band — how much load

The bands follow the developmental progression the framework already runs on — slow, at speed, under stress, sustained — because that is the ladder the gates climb.

- **D1 — Structured.** Moderate assignment, stable or predictable conditions, supervision close. The plan held. The member worked through structure.
- **D2 — Full tempo.** Moderate assignment at full operational tempo. Recurring situation types. No stress feature sustained.
- **D3 — Stressed.** At least one stress feature sustained across the assignment: real competition for resources, compressed decision windows, consequential trade-offs.
- **D4 — Design load.** Two or more stress features running at once, or genuine novelty — conditions outside the member's pattern base. This is the load expert skills are built for and decay without.

The features — what kind of load

Six, one letter each. A feature counts only when the member's own decisions had to answer it — not when it was merely present somewhere on the incident.

- **T — Tempo.** Decision windows compressed; the situation moved faster than the planning cycle.
- **R — Resources.** Orders unfilled, real competition, sharing across boundaries.
- **C — Consequence.** Life-safety territory live inside the member's span of authority.
- **N — Novelty.** Conditions the member's experience did not prepare them for.
- **S — Seams.** Coordination with neighboring elements as first-order work, not overhead.
- **I — Information.** The picture built on degraded, conflicting, or late information.

A code reads D3-TRS: band, then features.

Coding rules

- The member proposes the code at post-assignment reflection (F.4). The evaluator or supervisor confirms it. Disagreement resolves to the confirmer, with a note. Nobody codes their own record alone.
- The reflection has to show the feature doing work — a decision the feature forced. A claimed feature with no decision behind it comes off the code.
- Simulation is coded on the same scale, set by scenario design before the exercise runs. This is the fairness mechanism: a mild season or a quiet region cannot supply D3–D4, and simulation can.
- Gates read the log mechanically: the Level 2 gate needs D1 or higher; Level 3 needs D2 plus scored simulation; Level 4 needs D3; Level 5 needs a sustained D3–D4 record. An absence of D4 entries inside the review window is the decay signal — the gap goes in the record instead of being assumed away.

The log

One line per evaluated assignment or scored simulation:

Date	Incident / exercise	Position	Code	Feature basis (one line)	Confirmer

DIVS position anchors (worked example)

The bands, features, and rules above are defined once and inherited — the same pattern as the Section 6 shared skill areas. Each appendix supplies the position anchors: what each feature looks like in that job. The DIVS anchors:

Feature	What it looks like at division scale
T	Fire behavior or operational tempo outpacing the operational period plan; tactical decisions forced faster than the planning cycle can answer.
R	Orders unfilled; real competition with neighboring divisions; shared resources moving across the boundary mid-shift.

Feature	What it looks like at division scale
C	Life-safety territory live within the division: crews committed where escape routes, safety zones, or structures put lives inside the decision.
N	Fuels, fire behavior, tempo, or coordination structures outside the member's pattern base — the situation experience did not prepare them for.
S	Neighboring divisions, air operations, or shared boundaries as first-order work: dependencies that shape the division's own plan.
I	The division picture built on conflicting, degraded, or late information — intel that arrives after the decision it should have informed.

The Finance/Administration equivalent of T is a cost decision under a closing deadline, not a fire run. Each functional-area appendix writes its own anchor table with the Step 4 panel; the scheme itself does not change.

F.4 — Structured Reflection Template

After each assignment, the member writes a short record against their focus areas, following the DEBRIEF format the team already uses. An evaluator reviews it — the member never grades their own. The reviewer also checks the proposed difficulty code against the narrative: a claimed feature must appear as a decision the member had to make.

Header

Field	Entry
Assignment / exercise	
Position · dates	
Proposed code	Band-features, with one line of basis per feature claimed
Current focus areas	Carried from the last gate record (F.5)

Per focus area — three questions

- **1.** What did I try? The behavior worked on, and where in the assignment it was exercised.
- **2.** What happened? What the situation did in response — evidence, not evaluation.
- **3.** What do I now believe, and why? The updated read, with the reasoning stated so the reviewer can check it.

Standing entries — every reflection

- The conflict count for the assignment: how many cues I explained away, and what I did at the second one.
- The assumption that failed — or, if none did, the one that came closest.
- The read I changed — or the read I should have changed, and what kept me on it.

Reviewer block

- Code confirmed or adjusted, with the note if adjusted.
- Does the narrative support each claimed feature — a decision the feature forced? Features without decisions come off.
- Focus-area evidence noted for the gate record.
- Reviewer name and standing; forwarded to the exposure log and gate file.

F.5 — Gate Record

One record per gate decision. The form enforces three rules structurally: no gate passes on a single evidence stream; the evaluator's standing is marked, not assumed; and every result — pass or hold — leaves with focus areas and an exposure plan. A hold with a plan is development. A hold without one is just a rejection, and this form makes it impossible to issue one.

Section	Content
Member	Name · position · current level · gate sought
Evidence — field	Evaluated assignments: dates, codes from the exposure log, rating sheet references. The code must meet the gate's minimum band.
Evidence — simulation	Scored scenarios: IDs, designed codes, score sheet references.
Evidence — reflection	Reviewed reflections in the cycle: count and reviewer.
Three-stream check	All three streams present: Y / N. If N, the gate does not run.
Result	Pass · Hold
Focus areas — next cycle	Four or five, never more — chosen by the cost of leaving the gap open. The member should be able to say them from memory.
Exposure plan	The codes the next cycle needs, and where they will come from — field assignment or designed simulation. If the season cannot supply the band, the simulation venue is named here.
Review date	End of season, or after any evaluated assignment at or above target difficulty.
Evaluator	Name · holds the level being certified or higher: Y / N. Non-holder evaluations are marked in the record, not treated as equal.
Signatures	Member · evaluator. The member signs the focus areas and the plan, not just the result.

Appendix G — Evaluator Calibration Protocol

G.1 — Purpose and Standard

The pilot asks two questions. The behavior lists answer the first — do the behaviors tell people apart. This protocol answers the second — do different evaluators agree — with a number, not an impression. The primary measure is percent agreement per behavior item: on the same recorded performance, what fraction of evaluators gave the same rating. Agreement is reported item by item, because the item is what gets fixed. An item below the [panel] threshold is treated as an instrument defect first, not an

evaluator failure — the anchors get rewritten before anyone gets retrained. The pilot analysis may add chance-corrected statistics; the working number stays simple enough that a panel can act on it in the room.

G.2 — Session Design

- **Material.** Recorded scenarios from existing simulation courses. No new field burden; the venue already exists.
- **Independence first.** Evaluators rate independently first, using the F.1 and F.2 sheets, before any discussion. Independent ratings are the data; discussion afterward is the calibration.
- **Then the differences.** After independent rating, evaluators compare item by item and work out their differences. The session records which items drove disagreement and why — anchor ambiguity, unseen behavior, or genuine judgment difference. Only the first is fixed by rewriting; the record says which is which.
- **Difficulty coding on the same recordings.** Evaluators also code the same recordings with the F.3 scheme. Coding agreement is measured alongside performance agreement, item for item — band and features. If evaluators cannot agree on what an assignment demanded, gates cannot read the log mechanically, and the scheme's anchors get the same rewrite treatment as the behavior anchors.

G.3 — Cadence and Pool Rules

- **Pre-season, mandatory.** One session before the season. No one evaluates in a season they have not calibrated in.
- **Midseason, conditional.** A short session if the pilot data shows drift — agreement falling on items that previously held.
- **Non-holder marking (Section 7.3).** An evaluation done by someone who has not genuinely operated at the level being assessed is marked as such in the record — not treated as equal. Calibration does not upgrade a non-holder; it makes their ratings comparable.
- **Drift rule.** An evaluator persistently outside the group on repeated items recalibrates before their next evaluation counts. Persistent divergence after recalibration is a pool decision, not a paperwork one.
- **Professional safety.** Calibration records are never used against evaluators. Honest ratings must be professionally safe to give — an evaluator under pressure from promotion timelines or home units will drift toward friendly ratings, and the system will get the ratings it rewards instead of the ratings it needs. The calibration log exists to fix instruments and protect comparability, not to grade the graders.
- **Pool maintenance (from A.6).** The evaluator pool is maintained the way member skill is: evaluators return to operational assignments on a schedule, because evaluation judgment decays the same way operational judgment does.

G.4 — Records

One calibration log per session: date, scenarios used, evaluators present, per-item agreement on performance ratings, per-item agreement on difficulty coding, anchor changes ordered, and drift actions taken. The log feeds the pilot report directly — Phase 2’s second question is answered from this record and nowhere else.

References

Corpus works are published by Cogplexiti, Flagstaff, AZ. External works are cited in author-date form throughout.

- Cohen, Marvin S., Jared T. Freeman, and Bryan B. Thompson. 1995. *Training the Naturalistic Decision Maker*. Arlington, VA: Cognitive Technologies.
- Coil, Jayson. 2026. *Operational Leader’s Field Guide*. Flagstaff, AZ: Cogplexiti.
- Coil, Jayson. 2026. *The Incident Tier*. Flagstaff, AZ: Cogplexiti.
- Coil, Jayson. 2026. *The Second Skill*. Flagstaff, AZ: Cogplexiti.
- Coil, Jayson. 2026. *Sustaining Decision Performance*. Flagstaff, AZ: Cogplexiti.
- Dreyfus, Hubert L., and Stuart E. Dreyfus. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Klein, Gary A. 1993. “A Recognition-Primed Decision (RPD) Model of Rapid Decision Making.” In *Decision Making in Action: Models and Methods*, edited by Gary A. Klein, Judith Orasanu, Roberta Calderwood, and Caroline E. Zsombok. Norwood, NJ: Ablex.
- Klein, Gary A. 1998. *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.
- Smith, Patricia Cain, and Lorne M. Kendall. 1963. “Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales.” *Journal of Applied Psychology* 47 (2): 149–155.
- Tannenbaum, Scott I., and Christopher P. Cerasoli. 2013. “Do Team and Individual Debriefs Enhance Performance? A Meta-Analysis.” *Human Factors* 55 (1): 231–245.

“The difference between levels is the quality of the thinking, not the amount of time.”

A qualification tells you when a person is ready to start a job. It tells you nothing after that. The signature goes in the book, and from that day forward the system assumes the person is getting better just by doing the work — an assumption that fails wherever experience without structure builds confidence faster than it builds skill. Ask the system where any qualified member sits on the path from newly qualified to expert, what they should work on next, or whether ten years of experience is real growth or one year repeated ten times, and it cannot answer. For anyone.

Time in grade appears nowhere in this framework. Performance under load is the only thing that advances anyone.

The CIMT Skills Development Framework gives every Complex Incident Management Team position a five-level path from newly qualified to expert, built on the Dreyfus model and governed by one rule of movement: do it slow, then at speed, then at speed under stress, then at speed under stress consistently. Every gate draws on three evidence streams — field evaluation on difficulty-coded assignments, scored simulation, and structured reflection reviewed by someone else — and no gate passes on one stream alone. The framework arrives complete: a fully worked Division/Group Supervisor appendix, draft appendices for Planning, Logistics, Finance/Administration, and Command Staff, the full pilot instrument set, and the evaluator calibration protocol. It is a development tool, not a promotion board — every member carries four or five focus areas they can say from memory, and every gate result leaves with a plan. The drafts stay drafts until the pilot proves the instrument. Requirements follow measurement.

For training officers, incident commanders and command and general staff, geographic area training representatives, and the agency workforce leads accountable for what the bench can actually do.

The framework operationalizes the case made across the Cogplexiti Analysis line — the evaluation chapters of the Operational Leader’s Field Guide, The Second Skill, The Incident Tier, and Sustaining Decision Performance — into an instrument a workforce can run.

ABOUT THE AUTHOR

Jayson Coil is the founder of Cogplexiti LLC, an Assistant Chief with the Sedona Fire District, and a Type 1 Operations Section Chief on Southwest Incident Management Team 1. A U.S. Army veteran and red team instructor, he works at the intersection of operational command and the study of how organizations decide under uncertainty. He has helped develop the systems agencies use to manage major incidents and has presented for the National Fire Academy, the U.S. Forest Service, the Department of the Interior, FEMA, the Haas School of Business at UC Berkeley, and Los Alamos National Laboratory.

