Fixing the Field Training Bottleneck

The case for one connected foundation over a stack of separate tools

11.02.2025

Executive summary

Field training is the most important program most agencies run, and one of the least modernized.

It is the bridge between the academy and solo patrol. It is where an agency decides, in writing, whether a recruit is ready to work alone with a badge and a gun. That decision has to be fair, consistent, and defensible, because it will be questioned. By the recruit. By the union. Sometimes by a court.

Yet in a lot of agencies, this high stakes program still runs on paper forms, fading memory, and the private judgment of whichever officer happened to be assigned that day. Two field training officers watch the same behavior and score it differently. Documentation gets thin right when it matters most. Good recruits wash out and weak ones slip through, not because the standard was wrong, but because it was applied unevenly.

This paper explains why field training programs stall and what it takes to fix the core problem, which is inconsistent evaluation. The fix is not a new model. It is standardization done right: shared behavioral anchors, calibrated raters, structured daily documentation, and the ability to see patterns across your whole program instead of one recruit at a time.

The one program that shapes every officer who follows

Field training decides who becomes an officer. Everything downstream depends on it. Get it right, and you produce officers who are ready, agencies that can defend their probationary decisions, and a standard the whole department can trust. Get it wrong, and the cost compounds for years. A recruit who should have washed out becomes a problem officer. A strong recruit who was failed by an inconsistent evaluator is gone, along with the money spent recruiting and training them.

The stakes are legal as much as operational. When an agency removes a recruit from the program, that decision has to rest on a clear, documented record. The whole reason the modern field training model exists is defensibility. The original San Jose program, built in the 1970s and later adopted as a state standard and a benchmark for accreditation, was designed around one idea: break performance into specific behaviors, rate them on a consistent scale, and write it down every day. That structure is what lets an agency stand behind its decisions.

Structure on paper is not the same as structure in practice, though. And that gap is where programs stall.

Why good programs stall

Most FTO programs do not fail because the model is bad. They stall because the day to day execution wears down under real world pressure. A few forces do most of the damage.

Paper and memory. In many agencies, the Daily Observation Report is still a paper form or a document buried in a shared drive. At the end of a long shift, the field training officer fills it out from memory and whatever notes they scratched down between calls. Details fade. The report becomes a rough approximation of the day instead of an accurate record of it.

FTO burnout. Being a field training officer is a heavy add on to an already full job. You do your own police work, then you teach, correct, and grade someone else at the same time. Over months, that load causes frustration and burnout. Tired evaluators cut corners on documentation, which is exactly the documentation the agency will need later.

Staffing pressure. When an agency is short on officers, there is quiet pressure to move recruits through. Nobody says it out loud, but a program under staffing strain tends to pass people it might otherwise hold back. The standard bends without anyone deciding to bend it.

Inconsistent field training officers. FTOs are not trained evaluators by default. They are experienced officers, each with their own instincts about what "good" looks like. Without deliberate calibration, one FTO grades hard and another grades soft. The recruit's fate depends less on their performance than on who they drew.

No line of sight. Because the records live on paper or in scattered files, no one can easily see across the program. A sergeant cannot spot that one FTO fails everyone, or that recruits keep struggling with the same skill, or that a particular trainee is trending down across every phase. The signals exist, but they are buried in a filing cabinet.

Each of these is fixable. But they all point back to one root problem that deserves its own section.

The core problem: two officers, two different scores

Here is the failure at the heart of most struggling programs. It is not dramatic. It is quiet, and it happens every day.

One field training officer watches a recruit handle a call and rates the behavior a 2, unacceptable. Another FTO watches the same kind of behavior and rates it a 5, acceptable. Same conduct, opposite conclusion. The problem is not the recruit. It is that the two evaluators are not measuring against the same standard.

This is not a new observation. The people who built the modern field training model saw it clearly. The San Jose manual itself warns that standardization breaks down the moment one rater treats a behavior as unacceptable while another treats the identical behavior as fine. That gap between raters is the thing the whole system is supposed to close.

They closed it, originally, with hard work. The program's designers studied dozens of experienced field training officers and made them put into plain words the exact behavior behind each numerical score. What does a 4 actually look like? What separates a 3 from a 5? Those written anchors became the standardized evaluation guidelines. The point was to take the score out of the evaluator's gut and tie it to observable behavior that any trained FTO would rate the same way.

That is the whole game. A rating scale means nothing if two reasonable people apply it differently. Standardization is not about having a form. It is about having a form that different evaluators fill out the same way. Most stalled programs have the form. What they lack is the shared, enforced definition of what each score means, and any way to check whether their FTOs are actually applying it consistently.

What standardization actually requires

Fixing evaluation takes more than good intentions. It takes four things working together.

1. **Behavioral anchors for every rating.** Every score on the scale needs a plain language description of the behavior it represents, tied to the specific skill being rated. Not "communication: 4," but a written definition of what a 4 in communication looks like on a

real call. When the anchors are clear and shared, the score stops being an opinion and starts being an observation.

2. **Calibrated evaluators.** Anchors only work if FTOs are trained on them and checked against each other. Calibration means your field training officers watch the same performance and learn to score it the same way. It means a supervisor can see when one FTO drifts hard or soft and coach them back to the standard. Consistency is a skill, and it has to be built and maintained.

3. **Structured daily documentation.** The record has to be captured close to the event, not reconstructed from memory days later. It has to prompt the evaluator for the specific behaviors that justify the score, including the narrative that explains a low mark. When a recruit is not responding to training, that has to be documented at the time, with the remedial training that followed. Contemporaneous, structured records are what make a decision defensible.

4. **A single, consistent record.** Every DOR, every phase evaluation, and every remedial note for a recruit has to live in one place, in one format, from day one through the final decision. A record scattered across paper, email, and three FTOs' notebooks is not a record. It is a liability.

Do these four things and the core problem goes away. Two officers, measured against the same anchors and calibrated to each other, produce two scores that agree. That is standardization that actually holds.

Building a record that holds up

The reason to standardize is not neatness. It is defensibility.

When an agency decides a recruit is not ready, that decision has to survive scrutiny. The strongest defense is a complete, consistent, contemporaneous record that shows the recruit was evaluated fairly against a clear standard, was told where they fell short, was given real remedial training, and still did not meet the bar. That record cannot be assembled after the fact. It has to be built day by day, the right way, from the start.

This is also where the two big field training models meet their limits. The San Jose model is strong on documentation and standardization, which is exactly why it became the accreditation benchmark. Its critics argue it leans too hard on daily evaluation and not enough on how adults actually learn. The newer problem-based model, developed with federal support, answers that by emphasizing coaching, reflection, and critical thinking, but its lighter documentation worries agencies that need a paper trail for probationary decisions.

You do not have to pick a side to fix the bottleneck. The debate is about learning philosophy. The bottleneck is about execution. Whatever model you run, the recruit still needs consistent evaluation and the agency still needs a defensible record. Standardization serves both models. It is not a third camp. It is the foundation under all of them.

Seeing the whole program, not one recruit at a time

The last piece is the one paper-based programs can never deliver: a view across the entire program.

When every evaluation lives in one consistent system, patterns you could never see before come into focus. You can spot that one FTO fails recruits at twice the rate of everyone else, which is a calibration problem, not a recruit problem. You can see that trainees across the board struggle with the same skill in the same phase, which points at an academy gap or a program design flaw. You can catch a recruit trending down across phases early enough to intervene, instead of discovering it at the final evaluation.

This is the difference between managing field training and just running it. A program you can see is a program you can improve. This is the layer that a modern field training platform adds on top of standardized evaluation. Command HQ's field training application, COACH, is built around exactly this idea: capture the daily record in a consistent, structured, defensible form, and turn the accumulated data into a clear view of how recruits, evaluators, and the program as a whole are performing.

Where to start

You do not have to rebuild your program to fix the bottleneck. Start here.

- Write your anchors. For each rated skill, define in plain words what each score means in observable behavior. If your FTOs cannot point to the written definition of a 4, that is the first gap to close.
- Calibrate your FTOs. Get your evaluators scoring the same performance together until they agree. Make it a recurring practice, not a one time class.
- Move the record off paper. Capture the daily report in a structured form, at or near the time of the event, in one consistent place for every recruit.
- Look across the program. Once the data is consistent, start reviewing it by FTO, by skill, and by phase. Let the patterns tell you where to improve.

The bottom line

Field training is where an agency makes its most consequential personnel decisions, and too often it runs on tools that cannot support the weight. Programs do not stall because the model is wrong. They stall because evaluation is inconsistent, documentation is thin, and no one can see the whole picture.

The fix is standardization done seriously: shared behavioral anchors, calibrated evaluators, structured daily records, and a single consistent system that lets you see across the entire program. Get that right and the bottleneck clears. Recruits get a fair shot, decisions hold up, and the agency finally knows that the officer who reaches solo patrol earned it against a standard worth the name.