

Field Test Rubrics (Coach Only)

Quick pass/fail per item, one page per phase. The point isn't to nitpick, it's to confirm the skill is real. Send back anything that's missing or faked, that's part of the lesson too.

Your access codes and passwords are in the CODES.txt that came with your kit (or the one your Mission Config page generated). Keep this document and CODES.txt away from your operative. Everything else in the kit is fair game for them to find.

Phase 1 Field Test: Prove the Setup

Before you even open the evidence

- [] **Subject line = their codename + the word OPERATIVE.** If it's missing, they skimmed the file instead of reading it. Send it back and tell them to read the whole thing.

The six

1. **The Folders:** Finder shows `~/Documents/Business/` with all six subfolders, named correctly (01-Research, 02-Legal, 03-Brand, 04-Website, 05-Marketing, 06-Finances). - *Tell:* a phone photo of the screen instead of a real screenshot means they didn't use Cmd+Shift+4. Real screenshots are crisp, full-resolution, no glare or angle.
2. **The Loadout:** installed apps visible (Safari, Chrome, notes app, Claude, ChatGPT, password manager), plus a Google check: the 9-dot app-launcher grid or Drive open in the browser (Google isn't a download, so don't expect an app icon for it). - *Tell:* if half are missing, they haven't actually set up the toolkit.
3. **The Shortcut Reel:** an actual screen recording (a video file), not a photo or a description. Watch that they perform all five live: Spotlight launch, Cmd+Tab switch, Mission Control, window snap, screenshot. - *Tell:* if they "explain" the shortcuts in text instead of recording them, they may not actually know them. Make them record it.
4. **The Journal:** real entries, and a genuine "AI got it wrong" catch. - *Tell:* a filler entry like "the AI said $2+2=5$ " is a dodge. A real catch names something the AI actually told them that was wrong or out of date, and how they noticed.
5. **The Prompt Lab:** clear difference between the lazy and the sharp prompt, both AI answers present, the AI's self-critique present, and a learning sentence in their own words. - *Tell:* if the "lazy" and "pro" prompts are basically the same, they didn't get the point yet.

6. **The Explorer's Find:** a real shortcut/trick that wasn't on the Phase 1 list (e.g., Cmd+` to switch windows in an app, Cmd+Shift+T to reopen a closed tab, Cmd+L to jump to the address bar). - *Tell:* if it's one of the ones already listed, ask for one they actually discovered.

Verdict

- **All six clean:** Phase 1 verified. Pay the reward you agreed on.
- **Partial:** send back exactly what's missing. No full payout until it's complete. That's the gate doing its job.

This is also a live "AI being checked" moment for you: if their journal's AI-mistake entry is sharp, that habit is taking. That's the one to care about most.

Phase 2 Field Test: Prove the Idea

Before you even open the evidence

- [] **Subject line = their codename + OPERATIVE + PHASE 2.** Same rule as Phase 1. Wrong subject line means they skimmed. Back it goes.

The six

1. **The Idea Vault:** the full brainstorm list, 20 or more ideas, ugly ones included. - *Tell:* a tidy list of exactly ten "good" ideas means they cleaned it up after the fact. You want volume and a visible graveyard. Bad ideas on the page are evidence of real work, not sloppiness.
2. **The Scorecard:** five finalists scored on the rubric (interest, budget-realistic, time-realistic, real problem, laptop-friendly, 1-5 each), with the numbers visible. - *Tell:* if every row is 4s and 5s, they scored to justify a favorite instead of comparing honestly. You should be able to see why the winner won, and at least one finalist should have lost on the numbers.
3. **The Field Interviews:** proof of five real conversations with people in the target market: who (first name and how they fit), when, and the 2-3 things that surprised them. - *Tell:* five texts to friends who said "cool idea" don't count. If nothing surprised them, they weren't listening. At least one interview should have genuinely changed their thinking, ask them which one and watch how fast they answer.
4. **The Memo:** the 1-2 page decision memo: the one idea, why it won, and answers to the three validation questions (who specifically has the problem, what they do about it now, why they'd pay YOU). - *Tell:* a memo with no enemy is a pitch, not a decision. It should

name what almost killed the idea and why it survived anyway. Killing a good-sounding idea for a real reason counts FOR them, not against.

5. **The Customer:** one specific, picturable person. Age, situation, where they hang out online, what they've tried, what the problem costs them. - *Tell:* "everyone who likes X" is an automatic send-back. If you can't picture one real person from their description, neither can they.
6. **The AI Catch:** their Phase 2 "AI got it wrong" journal entry: the exact claim, what was actually true, and how they caught it. - *Tell:* idea-hunting season is prime time for AI confidently inventing market sizes, revenue numbers, and competitors. If they say "the AI was right about everything," push: did they actually verify, or did they stop checking?

Verdict

- **All six clean:** Phase 2 verified. Pay the reward you agreed on.
- **Partial:** send back exactly what's missing or thin. The two most common reships are homework-smelling interviews and a memo with no enemy.
- **Special case:** if they killed their leading idea for a real, evidence-backed reason and picked a different one (or asked to restart the hunt), that is a WIN. Verify the evidence, pay it like a pass, and say so out loud. Founders who can kill their own darling are rare.

The through-line, same as every phase: the AI catch. If it's sharp, the habit is building. If it's hollow, that's the thing to fix before Phase 3, not the scorecard math.

Phase 3 Field Test: Make It Real (Legal Foundation)

Before you even open the evidence

- [] **Subject line = codename + OPERATIVE + PHASE 3.** Same rule as always.

The six

1. **The Structure Memo:** compares all three options (sole prop, single-member LLC, S-corp-later) with the founder's *state-verified* costs, picks one, defends it. No real numbers = reship.
2. **The Name Check:** state business search + domain + social handles, screenshots of all three.
3. **The Filing:** confirmation present; if a parent signed because the founder is a minor, the founder prepared it.

4. **The Money Wall:** EIN + business account proven, with the numbers blacked out. Check they actually blacked them out; that's its own lesson in handling sensitive info.
5. **The Books:** ledger live with the computer logged as asset #1.
6. **The AI Catch:** documented: claim, truth, how caught. Legal and tax is where the AI is most confidently wrong; this is THE phase for it.

Auto-flags

Business money landing in a personal account (automatic reship, no exceptions); a memo that recommends an LLC "to be safe" without pricing the annual cost; tax facts quoted from the AI with no verification, that's backwards.

Live check

Ask "what do you owe in taxes and when?" If they can't answer in two sentences, the memo was decoration.

Phase 4 Field Test: Don't Blend In (Brand Foundation)

Before you even open the evidence

- [] **Subject line = codename + OPERATIVE + PHASE 4.**

The six

1. **The Brief:** one page with all four parts (what, who, difference, voice adjectives) plus 3 headlines + 3 captions written in the voice.
2. **The Recon:** 10 competitor screenshots with a same/different read.
3. **The Identity:** palette with hex codes + fonts, each with one line of reasoning. "It looked nice" isn't reasoning.
4. **The Mark:** logo as transparent PNG + square social version, WITH the iterations shown, AI first drafts next to the final.
5. **The Claim:** domain + handles secured and matching.
6. **The AI Catch:** documented. Image gen and brand advice miss constantly: unreadable logo text, "available" names that aren't, made-up design rules.

Auto-flags

First-draft logo shipped untouched; a “who it’s for” that describes everyone; handles that don’t match the domain.

Live check

Read one of their three headlines out loud, then ask “why does this sound like you and not like everyone else?” No answer = the voice isn’t real yet.

Phase 5 Field Test: Ship It Live (Website Build)

Before you even open the evidence

- [] **Subject line = codename + OPERATIVE + PHASE 5.**

The six

1. **The Live URL:** on their own domain. Open it yourself; don’t grade the screenshot.
2. **The Phone Walkthrough:** a phone screen-recording that walks every page.
3. **The Pages:** all five (home, about, product/service, contact) + footer legal links, no dead links.
4. **The Copy:** theirs, with draft-vs-final visible, and they can defend any paragraph live.
5. **The Pipe:** test transaction or captured lead proven, plus email live on the domain.
6. **The AI Catch:** documented. Web builds are full of it: platform features that don’t exist, settings menus that moved, code that “works” and doesn’t.

Auto-flags

Any lorem ipsum or placeholder (automatic reship); a preview/staging URL instead of the real domain; pasted code they can’t explain.

Live check

Pick one paragraph of site copy at random and ask why it says what it says. Then open the site on YOUR phone while they watch.

Phase 6 Field Test: Go Get Paid (Launch & First Traction)

Before you even open the evidence

- [] **Subject line = codename + OPERATIVE + PHASE 6.**

The six

1. **The Channels:** two platforms live + first 10 posts, mixed content, not ten of the same thing.
2. **The Outreach Log:** 20 named attempts with responses. DMs, emails, in-person count. Posting does not.
3. **The Paid Test:** ran at the size the budget dial allows, with coach-approved spend, and the founder can state what a click / lead / sale cost. (Near-free dial: the organic equivalent, same numbers read.)
4. **The Result:** first sale with receipts OR a no-sale write-up with numbers and changes. Both pass. Only spin fails.
5. **The Retro:** worked / didn't / change / next, plus feedback from 3 real humans.
6. **The AI Catch:** documented. Launch season special: invented benchmarks, "guaranteed" tactics, ad advice from three years ago.

Auto-flags

An outreach log that's really a posting log; a retro with adjectives but no numbers; revenue or benchmark claims sourced from the AI.

Live check

Ask "if you ran the launch again Monday, what's the first thing you'd do differently?" A founder who launched for real answers instantly. Then, whatever the answer, celebrate. They finished.