WORKSHOPSKILL · FACILITATION KIT

Training Evaluation Form — Facilitator Run Sheet

Folder: `training-evaluation-form` · Category: feedback forms · any group size · 7 min on the day + 3 min at day 30 · paper or online

License: CC BY 4.0

CONTENTS

What this form covers
Level 2: the self-report proxy
The day-30 follow-up
Variations
Agent version
Item-by-item interpretation
The day-30 email (template)
3 known failure modes of reading the results
Why eight reaction items
The keep/change questions
Known failure modes & rescues
Props in this folder
The 60-second read
Level 2: reading the confidence self-reports
The management one-pager (template)

7	Any	7 min on the day + 3 min at day 30
MINUTES	PEOPLE	SETTING

Training Evaluation Form — Facilitator Guide

A Kirkpatrick-ready evaluation pack for corporate training: eight reaction items, three confidence self-reports, two keep/change questions — plus the day-30 follow-up that turns feedback into behavior evidence. It produces a one-page report a manager can forward without translation.

What this form covers

Kirkpatrick's four levels, mapped honestly:

Level	What it asks	Where it lives in this folder
1 — Reaction	Did the room find it relevant, deep, well-run?	The 8 items on the form
2 — Learning	Can people now do the core things?	3 confidence self-reports (a proxy — see below)
3 — Behavior	Is it showing up in the work?	The day-30 follow-up (3 questions by email)
4 — Results	Did business numbers move?	**Not covered.** Not with a form. Be suspicious of any form that claims it.

The honest boundary at Level 4 is the feature: evaluation forms that pretend to measure business results produce numbers nobody believes, and then nobody believes the rest of the form either.

Why eight reaction items

Grouped, not piled: four content items (relevance, depth, instructor, materials), three delivery items (exercises, logistics, workload), one overall. In corporate training the delivery items matter as much as the content — a well-taught course with a broken sign-up flow gets scored like a bad course, and the item means will tell you which half to fix.

Level 2: the self-report proxy

The three confidence items are a proxy for learning — cheaper than a test, rougher than a test, and honest about being a proxy. Self-report inflates when the training felt inspiring; it inflates less when the items ask for narrow, checkable abilities ("explain it in two minutes", "apply it without help"). That is why these three are phrased narrow.

If you can collect a baseline — the same three items at sign-up — the delta becomes the credible number. If you cannot, report the absolute values and write "self-report, no baseline" in the report. That one line of honesty is worth more than a chart.

The keep/change questions

The two open questions are not comments; they are the agenda of the next design meeting:

- “What worked so well it should stay exactly as it is?” — protects the good parts from a redesign that “improves” them away.
- “What should change in the next run?” — the fix list, in the participants’ own words.

For repeat programs this pair is worth more than every rating combined: ratings tell you where you stand, keep/change tells you what to do.

The day-30 follow-up

Thirty days after the training, send three questions by email (template in the debrief sheet): used-it frequency, what helped or blocked, what support would make it stick. This is the Level 3 behavior check — the level most evaluation never reaches, and the level that answers the question the sponsor actually asked: “did anything change?”.

Expect a third of the response rate of the in-room form. That is normal and still enough: behavior counts from a 30% sample are evidence; satisfaction means from a 90% sample are not evidence of behavior at all.

Known failure modes & rescues

1. **The form becomes a satisfaction contest between vendors.** The moment item means are used to rank instructors, answers turn diplomatic and the instrument dies. Rescue: report room-level means to the program owner, never per-instructor league tables.
2. **The day-30 email never gets sent.** The pack halves itself: Level 3 evidence requires the follow-up, and it is always the first thing a busy calendar drops. Rescue: schedule the follow-up email the same day you run the training — write it before you forget who was in the room.
3. **Self-reports reported as learning gains.** Without a baseline, a 4.3 confidence mean is a feeling, not a gain. Rescue: say “self-report, no baseline” in the report — one honest line beats a decorated chart.

Variations

- **Multi-session programs:** run the form after each session; run the day-30 follow-up once, 30 days after the last session.
- **Compliance documentation (Taiwan / Hong Kong):** the item means plus the follow-up behavior counts are exactly what quality systems want as training-effectiveness evidence (TTQS course documentation, QF-style records). File the room-level summary; individual sheets stay private.
- **Vendor-delivered training:** add one item — “My manager knew what I was learning” — the cheapest measure of the transfer climate you will ever install.

Props in this folder

- `training-evaluation-form-prop-cards.pdf` — poker print-ready deck: instruction card, reaction-item cards, confidence card, keep/change card
- `training-evaluation-form-guide.md` — this guide
- `training-evaluation-form-debrief.md` — reading the results, the day-30 email template, the management one-pager
- `training-evaluation-form-assessment.json` — the full question set as plain JSON
- `training-evaluation-form-paper-form.pdf` — the A4 answer sheet: the measure as a print-and-fill form
- `training-evaluation-form-workbook.pdf` — participant workbook: the idea in plain words, guided items, notes and action pages
- `training-evaluation-form-slide-deck.pdf` — the slide deck for teaching or projecting
- `training-evaluation-form-slide-deck.html` — the same deck for any browser, speaker notes on N
- `training-evaluation-form-license.md` — what you may do with all of this, in plain language
- `training-evaluation-form-activity-poster.pdf` — The whole activity on one wall (A3): meta, run of show, debrief questions, watch-outs.
- `training-evaluation-form-start-here.md` — one page, thirty seconds, then print and go
- `SKILL.md` — machine-readable manifest for AI agents
- `training-evaluation-form-run-sheet.pdf` — the print edition: this run sheet + the debrief sheet, bound as one A4 PDF

Agent version

Every file in this folder has a stable public URL — an AI agent can fetch the cards, the reading guide and this run sheet with three curl commands (see `SKILL.md`), or build the online evaluation in one `formlm-cli smart generate` step — including the day-30 follow-up as a scheduled second collection. The paper version needs no screen: cards, pens, and an email client you already own.

Training Evaluation Form — Reading the Results

> One rule first: reaction items describe the day; only the day-30 answers describe the work. Report them separately, never averaged together.
>
> Use with `training-evaluation-form-guide.md`. Reading takes 20 minutes for a cohort of 20.

The 60-second read

1. **Overall item at 4.0 or higher, but relevance under 3.5?** The training was enjoyable and aimed at the wrong work. This is the most common corporate failure and the item means catch it precisely.
2. **Delivery items vs content items.** Logistics and workload under 3.0 while content rides high: fix the program administration, not the course.

3. **Day-30 behavior count above zero?** The training is alive. Zero uses in 30 days — the course was an event, not a change, and no reaction score repairs that.

Item-by-item interpretation

Item	Construct	Low (≤ 2.5) means	First action
1	Relevance	Content missed the actual work	Rewrite the needs intake before the next run
2	Depth	Correct generalities, no depth	Replace one lecture with one worked case
3	Instructor	Questions met with slides	Instructor prep session on Q&A handling
4	Materials	Handouts filed and forgotten	Cut to one page people will reopen
5	Exercises	Listening outweighed doing	Add one exercise per half-day
6	Logistics	Sign-up, schedule or room friction	Fix the admin flow — it is dragging the scores
7	Workload	Course load collided with the job	Reschedule or split the cohort
8	Overall	The room's honest verdict	Read items 1-7 before reacting to this one

Level 2: reading the confidence self-reports

The three items (apply without help / explain in two minutes / know where to look when stuck) are deliberate narrow abilities. Read them as a set:

- All three low — the course taught awareness, not capability. Not a failure; but call it awareness training in the report.
- "Apply without help" low, the other two high — people understood and remember, but the doing needs support: schedule a practice session, not a repeat of the course.
- With a baseline: the delta is your Level 2 number. Without: report absolutes plus the words "self-report, no baseline".

The day-30 email (template)

Send 30 days after the last session, from the person the participants know, subject: "30 seconds — did any of it stick?":

> You attended [program name] on [date]. Three questions, 30 seconds, anonymous:
>
> 1. Since the training, I have used the main method: (never / once / 2-5 times / weekly) — circle one
>
> 2. What helped or blocked the using? — one sentence
>
> 3. What support would make it stick? — one sentence
>
> Nothing else. No follow-up will chase you.

The last line is not politeness; it protects the response rate of the next program.

The management one-pager (template)

> **[Program name] — evaluation summary, [date], N=[n], response [r]%**
>
> - Reaction: overall [x.x]/5; lowest item [name] [x.x]; highest item [name] [x.x]
> - Confidence (self-report[+ delta]): apply [x.x][(Δ+x.x])], explain [x.x][(Δ+x.x])], recover [x.x][(Δ+x.x])]
> - Day-30: [n] of [n] replied; used-it count [n] (never [n] / once [n] / 2-5× [n] / weekly [n])
> - Keep: [one verbatim, permissioned] — Change: [one verbatim, permissioned]
> - Recommended next run: [one sentence, from the keep/change list]

One page. If it does not fit on one page, the program does not fit in a manager's attention — cut until it does.

---

3 known failure modes of reading the results

1. **Levels averaged together.** A 4.5 reaction mean next to a zero behavior count is a failed program with a nice cover photo. Rescue: report levels in separate lines — the gap **is** the finding.
2. **The one-pager grows a second page.** Charts, percentile bands, benchmark logos. Rescue: the template above is the ceiling; anything past it goes in an appendix nobody reads.
3. **The day-30 obstacles go unread.** The follow-up's second question (what blocked the using) is free transfer-climate consulting. Rescue: read those sentences first — they are the program's actual to-do list.