WORKSHOPSKILL · FACILITATION KIT

培训评估表（Training Evaluation Form） — 引导者执行手册

文件夹： `training-evaluation-form` · 类目：反馈表单 · 任意人数 · 当场 7 分钟 + 第 30 天 3 分钟 · 纸面或在线

许可：CC BY 4.0

目录

这份表覆盖什么
第 2 级：自评代理指标
第 30 天追踪
变体
代理版
逐题解读
第 30 天邮件（模板）
读结果的 3 个已知翻车点
为什么是八道反应题
保持 / 修改两题
已知翻车点与救援
文件夹里有什么
60 秒速读
第 2 级：自信度自评怎么读
管理层单页报告（模板）

7	不限	当场 7 分钟 + 第 30 天 3 分钟
分钟	人数	形式

培训评估表——使用指南

企业培训的柯氏（Kirkpatrick）级评估包：八道反应题、三道自信度自评、两道保持/修改题——外加第 30 天的追踪邮件，把反馈变成行为证据。最后产出一份管理者无需翻译就能直接转发的单页报告。

这份表覆盖什么

柯氏四级，如实对应：

级别	问的是什么	在文件夹里的位置
1 —— 反应	学员觉得内容相关、有深度、组织得好吗?	表上的 8 道题
2 —— 学习	大家现在会做核心的事了吗?	3 道自信度自评（代理指标——见下）
3 —— 行为	它出现在工作里了吗?	第 30 天追踪（邮件 3 题）
4 —— 结果	业务数字动了吗?	**不覆盖。** 表格测不了。谁声称能测，就警惕谁。

守住第 4 级这条诚实边界，正是这份表的特性：假装能测业务结果的评估表，产出的数字没人信，然后整张表都没人信了。

为什么是八道反应题

分组而不是堆砌：四道内容题（相关性、深度、讲师、材料）、三道交付题（练习、行政、工作量）、一道总体。企业培训里交付题和内容题一样重要——一门讲得好的课配上报名流程一团乱，会被打出差课的分数，而逐题均值会告诉你该修哪一半。

第 2 级：自评代理指标

三道自信度题是学习的代理指标——比测验便宜，也比测验粗糙，而且老实承认自己只是个代理。课上得让人激动时自评会膨胀；题目越窄、越可核对（「两分钟内讲清楚」「不用帮忙就能上手」），膨胀越小。这三道就是这么写的。

如果能收到基线——报名时收同样三道题——差值就成了可信的数字。收不到就报绝对值，并在报告里写明「自评、无基线」。一行诚实的字，比一张图表值钱。

保持 / 修改两题

两道开放题不是留言区；它们是下一次设计会议的议程：

- 「什么效果好到应该原样保留？」——保护好的部分不被下一次「改进」改没了。
- 「下一场该改什么？」——修改清单，用参与者自己的话写。

对重复开办的课程，这一对题比全部评分加起来都值钱：评分告诉你现状，保持/修改告诉你做什么。

第 30 天追踪

培训后 30 天，用邮件发三道题（模板在解读表里）：使用频率、什么帮了或拦了、什么支持能让它留下来。这是第 3 级行为检查——多数评估到不了的级别，也是回答买单方那个真正问题（「有什么变了吗？」）的级别。

预期回收率是在场表的三分之一。这是正常的，也够用：30% 样本的行为计数是证据；90% 样本的满意度均值根本不是行为的证据。

已知翻车点与救援

1. **评估表变成供应商满意度竞赛。** 题目均值一旦被拿来给讲师排座次，答案集体变外交辞令，工具就死了。救援：只向项目负责人汇报全场层面均值，永远不做讲师排行榜。
2. **第 30 天邮件永远没发。** 评估包等于自废一半：第 3 级证据全靠追踪，而它永远是繁忙日程里第一个被放弃的。救援：办培训当天就排好 30 天后的邮件——趁你还记得屋里坐的是谁，先把它写好。
3. **自评被当成学习增量汇报。** 没有基线时，4.3 的自信度均值是一种感觉，不是增量。救援：在报告里写明「自评、无基线」——一行诚实的字胜过一张花哨的图。

变体

- **多场次项目：** 每场结束都收一次表；第 30 天追踪只做一次——最后一场后的 30 天。
- **合规文档（台湾 / 香港）：** 题目均值加追踪行为计数，正是品质系统要的培训成效佐证（TTQS 课程问卷佐证、资历架构一类的记录）。归档全场汇总；个人答卷保持私密。
- **供应商交付的培训：** 加一道题——「我的直属上级知道我在学什么」——这是你能做的最省事的迁移氛围测量。

文件夹里有什么

- `道具卡.pdf` ——扑克规格印刷卡组：说明卡、反应题卡、自信度卡、保持/修改卡
- `执行指南.md` ——本指南
- `复盘问题.md` ——怎么读结果、第 30 天邮件模板、管理层单页报告
- `测评预设.json` ——完整题目集，纯 JSON
- `答卷.pdf` ——A4 答卷：量表的打印填写版
- `学员工作簿.pdf` ——学员工作簿：通俗讲解、带引导的题目、笔记与行动页
- `演示文稿.pdf` ——讲课或投屏用的演示文稿
- `演示文稿.html` ——同一套演示文稿的浏览器版，N 键讲者提示
- `使用许可.md` ——一句话说明这套材料你能做什么
- `活动海报.pdf` ——一面墙讲完整场活动（A3）：速览、流程、复盘问题、注意事项。
- `快速上手.md` ——一页，三十秒，然后打印进场
- `SKILL.md` ——供 AI 代理读取的机器可读清单

- `运行手册.pdf` ——印刷版：本执行脚本 + 复盘表，合订成一份 A4 PDF

代理版

本文件夹的每个文件都有稳定的公开 URL——AI 代理用三条 curl 命令就能拿到卡组、解读指南和本指南（见 `SKILL.md`），或用一条 `formlm-cli smart generate` 命令建好在线评估——包括把第 30 天追踪排成第二次回收。纸面版完全不需要屏幕：卡片、笔，加上你本来就在用的邮箱。

培训评估表——怎么读结果

> 先立一条规则：反应题描述那一天；只有第 30 天的答案描述工作。分开汇报，永远不混在一起求平均。
>
> 配合 `执行指南.md` 使用。20 人一班，读结果约 20 分钟。

60 秒速读

1. **总体题 ≥ 4.0，但相关性低于 3.5？** 培训本身讨喜，可惜嫖错了工作。这是企业培训最常见的失败，逐题均值能精确抓住它。
2. **交付题 vs 内容题。** 行政和工作量低于 3.0、内容题却很高：修的是项目管理，不是课程。
3. **第 30 天行为计数大于零吗？** 培训还活着。30 天零使用——课程是一个事件，不是一次改变，任何反应分数都救不回来。

逐题解读

题号	构念	低分（≤ 2.5）意味着	第一步动作
1	相关性	内容没对上实际工作	下一场之前重做需求调研
2	深度	正确的泛论，没有深度	把一场讲座换成一个完整案例
3	讲师	提问被幻灯片挡回去	给讲师做一次答疑应对的备课
4	材料	讲义发完即忘	砍成一页大家会再打开的东西
5	练习	听的时间多过做的时间	每半天加一个练习
6	行政	报名、日程或场地有摩擦	修行政流程——它在拖累分数
7	工作量	课程负载撞上了本职	改期或拆班
8	总体	全场的诚实判词	先读 1–7 题再对这题做反应

第 2 级：自信度自评怎么读

三道题（不用帮忙就能上手 / 两分钟讲清楚 / 卡住时知道去哪查）刻意选的都是窄能力。当成一组读：

- 三道全低——课程教了认知，没教能力。不是失败；但报告里要如实叫它认知培训。
- 「不用帮忙就能上手」低、另两道高——大家听懂了也记得，但上手需要支持：排一次练习场，不是把课再上一遍。
- 有基线：差值就是你的第 2 级数字。没有：报绝对值，加上「自评、无基线」四个字。

第 30 天邮件（模板）

最后一场后 30 天发，用参与者认识的那个人的邮箱，主题：「30 秒——有什么留下来了吗？」：

> 你在 [日期] 参加了 [项目名]。三个问题，30 秒，匿名：
>
> 1. 培训之后，我把主要方法用了：（没用过 / 用过一次 / 2–5 次 / 每周都在用）——圈一个
> 2. 什么帮了或拦住了使用？——一句话
> 3. 什么支持能让它留下来？——一句话
>
> 就这些。不会有后续来催你。

最后一句不是客气；它保住的是下一个项目的回收率。

管理层单页报告（模板）

> **[项目名] —— 评估摘要，[日期]，N=[n]，回收 [r]%**
>
> – 反应：总体 [x.x]/5；最低题 [题名] [x.x]；最高题 [题名] [x.x]
>
> – 自信度（自评[+ 差值]）：上手 [x.x][（Δ+x.x）]、讲清 [x.x][（Δ+x.x）]、自救 [x.x][（Δ+x.x）]
>
> – 第 30 天：[n]/[n] 回复；使用计数 [n]（没用过 [n] / 一次 [n] / 2–5 次 [n] / 每周 [n]）
>
> – 保持：[一句授权原话] —— 修改：[一句授权原话]
>
> – 下一场建议：[一句话，来自保持/修改清单]

一页。放不进一页，就说明项目放不进管理者的注意力——删到放进为止。

读结果的 3 个已知翻车点

1. **把各级别混起来求平均。** 4.5 的反应均值旁边一个零行为计数，是一个带漂亮封面照的失败项目。救援：分级分行汇报——差距本身**就是**发现。
2. **单页报告长出第二页。** 图表、百分位带、对标 logo。救援：上面的模板就是上限；超出的部分进没人读的附录。
3. **第 30 天的障碍没人读。** 追踪的第二问（什么拦住了使用）是免费的迁移氛围咨询。救援：先读那些句子——它们才是项目真正的待办清单。