



信息通告

中国民用航空局飞行标准司

编 号：IB-FS-OPS-006

下发日期：2022 年 8 月 23 日

循证训练（EBT）评分系统建设指南

目 录

1 目的和依据	1
2 参考资料	1
3 评分系统的原则	1
4 评分系统的目标	2
5 评分系统的技术规范	3
5.1 一般原则	3
5.2 计分制	3
5.3 VENN 模型	3
5.4 结果纬度	5
5.5 颗粒度	6
5.6 未展现的胜任力	6
6 教员间信度（评分一致性）	6
7 数据收集要求	7
7.1 评分指标层级	7
7.2 各模块使用的最低评分层级要求	7
8 针对性训练和补充训练建议	7
8.1 针对性训练建议	7
8.2 补充训练建议	8
9 生效和废止	8
附件 1：评分对于 TEM OUTCOME 安全裕度下降的补充举例说明 ...	9
附件 2：结果维度的评估实施样例	10
附件 3：EBT 新教员常见评估失准情况分析	13

1 目的和依据

为落实《关于全面深化运输航空公司飞行训练改革的指导意见》精神，统一循证训练（以下简称 EBT）实施过程中的评估标准，确保 EBT 评分系统的客观性、一致性和可操作性，依据《中国民航运输航空飞行员技能全生命周期管理体系（PLM）建设实施路线图》，制定本信息通告。本信息通告为航空运营人和航空训练机构（以下简称运营人和训练机构）建设 EBT 评分系统提供指导性建议。

2 参考资料

《EBT 实施指南》（国际航空运输协会）

《飞行员、教员、检查员胜任力评估指导材料》（国际航空运输协会）

《Recurrent training and checking》（EASA ORO.FC.230）

《Evidence-based training》（EASA ORO.FC.231）

《EBT programme assessment and training topics》（EASA ORO.FC.232）

3 评分系统的原则

EBT 评分系统应当具备以下九项特性：

（a）公平性和准确性

EBT 评分系统应当输出客观公平、有针对性的评价，并且其结果是可靠的、准确的、一致的。同时应当避免各类主观偏见，例如光环效应、教员或检查员的倦怠、机械式“勾选”、正面或者负面的偏见。

（b）明确性

EBT 评分系统应当在不受个体主观解释或混淆的影响下，输出透明、清晰、完整的评价。此系统还应当解决飞行员在某些条件下没有机会展示其特定胜任力的问题。

（c）实用性

EBT 评分系统应当简单易用、便于理解、利于管理，能有效防止无意差错。

(d) 合规性

评分系统应当持续符合局方规范性文件、运营人和训练机构运行/训练政策的相关要求，并长期保存相关记录，确保记录可追溯。

(e) 持续性

评分系统应当提供足够数量和高质量的证据，以便持续改进培训系统，改善受训人员的绩效表现。一方面，该系统可通过分数识别趋势，帮助分析和解决存量、增量或者潜在问题，进而优化培训系统。另一方面，它能够反馈对受训人员的评估情况，以帮助改善评分的一致性并且优化评分系统。

(f) 激励性

评分系统应当体现激励性、可信性、尊重性，易于让人接受，确保教员和学员不会对评分产生畏惧心理。

(g) 数据化

评分系统应当依托数据管理技术，提供安全可控、便捷输出、易于记录的高质量电子数据，同时兼容不同的媒介、分析和演示工具。

(h) 通用性

评分系统应灵活适用各类场景，能够广泛应用于运行和训练。

(i) 警觉性

评分系统应当具备对无效评分的警觉性，以避免评分过程中存在的偏差，能较好地统一教员间评分的内部标准。

4 评分系统的目标

评分系统应当具有明确的目标，同时满足使用者的实践需求。对于局方，该系统为颁发和更新飞行员执照/等级的评估工作提供支持。对于运营人和训练机构，该系统可衡量个人、机组、机队（仅对运营人）的表现，以及确定个人和机队（仅对运营人）的提升需求，确保

执行航线飞行任务的飞行员满足能力表现期望。对于飞行机组成员，该系统提供培训期间的评估信息，以促进机组提高胜任力。

5 评分系统的技术规范

5.1 一般原则

评估是贯穿培训阶段的持续观察过程。该过程应当按照四个步骤有序实施，即观察、记录、分类和评估。这是一个循序渐进的过程，其目的是防止主观过快地得出不正确的结论。

5.2 计分制

EBT 评分系统采用 5 分制记分方法：

- (a) 1 分表示不合格（不能回归航线运行）；
- (b) 2 分表示最低可接受（可以回归航线运行的最低要求）；
- (c) 3 分表示中等（机队的平均水平）；
- (d) 4 分和 5 分表示表现高于平均水平（其中 4 分为优秀，5 分为典范）。

5.3 VENN 模型

5.3.1 文字图片（Word-picture）表达形式可以确保准确体现相关标准，特别对于评分时的公平性、准确性和清晰度。分数可以用标准化的文字图片来描述，如下表：

1 分 - 不合格	飞行员很少在需要时展示任何相关行为指标，未能展示有效的该项胜任力，从而导致不安全的情况。
2 分 - 最低可接受	飞行员只是有时在需要时展示一些行为指标，仅展示了最低可接受水平的该项胜任力，总体而言并未导致不安全情况。
3 分 - 中等	飞行员时常在需要时展示许多行为指标，展示了足够的该项胜任力，从而确保了安全运行。

4 分 - 优良	飞行员经常需要在需要时展示大部分行为指标，展示了有效的该项胜任力，从而提高了安全性。
5 分 - 典范	飞行员始终需要在需要时展示所有行为指标，展示了典范的该项胜任力，从而显著提高了安全性和效率。

5.3.2 VENN 模型具有三个维度：

- (a) HOW OFTEN = 频次（例如：有时演示……）；
- (b) HOW MANY = 数量（例如：展示许多指标……）；
- (c) OUTCOME = 结果（例如：由于沟通不及时……导致了一个不安全的情况）。

注：胜任力评估维度中的程度（HOW WELL）是由相关胜任力的行为指标（OB）所展示的数量、频率以及威胁/差错管理的相应结果组成。

评分图例如下：

行为指标		TEM 的结果	胜任力评估	分数
数量	频次			
很少	几乎没有	不安全	不合格	1
一些	有时	安全裕度下降	最低可接受	2
许多	时常	安全	中等	3
大多数	经常	安全*	优良	4
全部	总是	提升安全	典范	5

三列中最低的一栏

安全*：主动的安全水平，提高或有可能提高安全裕度

5.3.3 分数的确定标准

为了确保一致性，5分制的量化管理应当遵循以下标准：

- (a) 1 分 - “很少，几乎没有”，仅仅超过 0%;
- (b) 2 分 - “一些”，接近 50%;
- (c) 3 分 - “许多” 在一些和大多数之间;
- (d) 4 分 - “大多数” 作为绝大多数，差不多为 90%;
- (e) 5 分 - “全部，几乎全部”，100%或接近 100%。

注：2 分表示最低可接受，即展示出约为所需的 50%的指标。因此，评估人员应当避免简单的划分，如考虑使用 100%除以 5，用以对应五个等级（即避免 1 级为 0%至 20%，2 级为 20%至 40%，等等）。对仅为 21%的 2 级（合格）进行分类是不合适的，经验丰富的教员应当以更全面的方式来观察学员的行为并且评估其能力，而不应机械的套用数学公式。

5.4 结果纬度

5.4.1 首先要明确与 TEM 结果相关的行为所对应的胜任力。教员在考虑结果所对应的分值时，不应简单的使用事故征候或差错的标准来评判，而应当关注学员在整个过程当中表现出的胜任力程度。对于结果维度又分成两个方面，一是对威胁/差错管理的数量和及时性，二是“管理后的结果”对安全裕度的影响。

注：评分对于 TEM OUTCOME 安全裕度下降的补充举例说明见附件 1，局方将不定期在“飞行人员信息咨询”网站（<http://pilot.caac.gov.cn>）上发布 EBT 评分样例，包括视频和文字材料。

5.4.2 判断标准和方法

(a) 1 分 - 没有适时管理所有的威胁/差错，导致一个不可接受的安全裕度下降;

(b) 2 分 - 在一些场景中没有管理好威胁/差错，导致一个有限的或者临时的安全裕度降低;

(c) 3 分 - 对于许多预期的威胁有准备、有缓解，对于非预期的威胁能识别和有效应对，并能适时管理差错，保持了安全裕度;

(d) 4 分 - 对于大多数预期的威胁有准备、有缓解，对于非预

期的威胁能识别、能缓解，并能迅速管理差错，提升了安全裕度；

(e) 5 分 – 对于所有预期的威胁有准备、有缓解，对于非预期的威胁能识别、能缓解，并能立即管理差错，大大提升了安全裕度。

注 1: 结果维度的评估实施样例，见附件 2。

注 2: EBT 新教员常见评估失准情况分析，见附件 3。

5.5 颗粒度

为了更有效的收集评估信息，在给予 1-5 的评分同时，应当标注出显著的或不足的行为指标；对于关键胜任力，应当记录机组在应对威胁/差错时的具体行为。

5.6 未展现的胜任力

5.6.1 首先由课程开发人员确定在本场景中无需展现的胜任力，教员/检查员无需评估此胜任力，直接勾选“未观察到（not observed）”，不进行评分。

5.6.2 首先对于课程实施过程中由于机组之间配合的问题，比如经验丰富的机长和一个新副驾驶搭组（新副驾驶很多动作行为都由机长接管，没有机会得以展现），教员在评分系统中勾选“未观察到（not observed）”的同时，需要说明相关情况，并根据情况评分或不评分。如机长直接接管操纵，则无需对 FPM 评分，可通过飞后讲评确定机组程序和知识类胜任力情况后，再进行评分。

6 教员间信度（评分一致性）

6.1 运营人和训练机构应当建立教员间信度的优化体系。该体系应当基于多维数据，核实教员的评分准确性和一致性，并识别出一致性较弱的领域，以改进评分系统的质量和有效性。

6.2 运营人和训练机构应当制定相应程序，对不符合间信度要求的教员提出改进建议，明确绩效标准，采取提升行动。

6.3 运营人和训练机构应当保留实施 EBT 课程的合格 EBT 教员清

单，并持续更新。

7 数据收集要求

7.1 评分指标层级

收集的数据应当符合以下指标层级的要求：

(a) 0 级评分指标（合格指标）：数据指标提供飞行员是否合格的信息；

(b) 1 级评分指标（能力指标）：来自评分系统的可量化的数据——胜任力的数字等级（如：1 至 5）；

(c) 2 级评分指标（可观察行为指标）：教员记录课程中正向/负向的 OB；

(d) 3 级评分指标（TEM 指标）：教员记录威胁、差错或者降低安全裕度的行为。

7.2 各模块使用的最低评分层级要求

(a) 机动检查：0 级，当结论为不合格时应当适用 2 级；

(b) 航线评估模块：1 级；

(c) 基于场景的训练（SBT）：1 级。

注 1：任何胜任力分数为 2 或 1，都应当适用 2 级评分层级。

注 2：不同的 EBT 实施阶段对应不同的评分指标层级，应当按照局方批准 EBT 方案，采用相应的评分指标层级。

8 针对性训练和补充训练建议

8.1 针对性训练建议

前一次复训中某一胜任力为 2 分，在本次复训航线评估中该胜任力再次为 2 分，本次复训的训练课应当采用针对性训练。以下述情况为例：

(a) 上一次复训的第三节课的某胜任力评分为 2 分，本次复训的第一节航线评估该项胜任力再次为 2 分。

(b) 上一次复训的第一节航线评估的某胜任力评分为 2 分，第二次复训的第一节航线评估该项胜任力再次为 2 分。

8.2 补充训练建议

(a) 连续三次复训中，同一胜任力为 2 分，机组可以进行航线运行，但应当在第三次复训结束后的 3 个月内完成一次模拟机补充训练。

(b) 同一次复训中如果在最后一节训练课有多于 4 个胜任力评为 2 分，机组可以进行航线运行，但应当在 3 个月内完成一次模拟机补充训练。

注 1: **针对性训练**是指为飞行员个人提供的“量体裁衣”式训练，可能需要运营人和训练机构开发相应的教学大纲以供教员参考，可以结合 SBT 完成，不需要增加额外模拟机时间，但可能增加 CBT（基于计算机的训练）或者额外的讲评。

注 2: **补充训练**是指针对性课程以外，额外增加的模拟机训练课程。

9 生效和废止

本信息通告自下发之日起生效。

附件 1：评分对于 TEM OUTCOME 安全裕度下降的补充举例说明

1. 起飞后发动机火警，如果机组因为报 MAYDAY 或其他沟通原因延误执行灭火程序，我们认为降低了安全裕度，相关胜任力不能为 3 分或以上，但是如果一直到落地都没有执行灭火一定为 1 分，如果在起飞和爬升阶段以外才处理灭火程序，则在 1-2 分之间，需要与其他场景中的该项胜任力结合进行评分。

2. 模拟机 TCAS 场景，在空中机组观察在前方有一架相对飞机正在爬升，且有可能穿越自己的高度，准备提醒管制员是否自己需要改变高度或航线，在沟通的时间段内出现了 RA 警告，RA 指令向下增加下降率，而机组观察冲突飞机是从下而上穿越自己高度，所以决定不按 RA 指令向上拉起飞机，两秒后冲突解除。虽然没有引起后果，但违背了 TCAS 的处置原则，没有产生后果是因为模拟机的设置限制，所以这个关键胜任力应该为 1 分，而且很有可能涉及 PRO 胜任力。

3. 对于不合规的情况：

(a) 对于进近阶段，在没有提升安全裕度的前提下，未按照 SOP 的要求设置着陆形态，该相关胜任力为 2-3 分。

(b) 未执行检查单，或者执行检查单时未落实着陆形态设置要求，该相关胜任力为 1-2 分。

(c) 在最后进近阶段，在机载设备警告的情况下，未执行复飞，该相关胜任力为 1 分。

(d) 对于落地长平飘，在没有提升安全裕度的前提下，如果在着陆区落地，该相关胜任力为 3 分；如果在着陆区以外落地（较长跑道）该相关胜任力为 2-3 分；如果在着陆区以外落地（短跑道），该相关胜任力只能为 1-2 分。

附件 2：结果维度的评估实施样例

评分应当基于整个场景，而不仅仅是贯彻一个特定的场景元素（例如一个飞行阶段）。以下提供一个基于威胁/差错管理（TEM）模型的示例。教员可以先参照该模型，按照文字图片所示进行评分。该模型侧重于确定 1 分或者 2 分（确定 1 分和 2 分是 EBT 评分的关键），以及 2 分以上的等级，但对于确定实际评分为 3 分、4 分或者 5 分的评分存在无效的可能性。

1. 对于无意不合规行为的评分（例如，机组未发现或者未及时纠正错误，但未导致不安全结果）。有两种类型：

（a）无意不合规，没有产生后果。

例如，PF 高度调置错误，而 PM 正忙于其他事务。当错误发生时，由于 PF 和 PM 之间没有确认高度，且未与 ATC 之间再次确认高度，导致机组未能识别该差错。但在突破高度之前，机组间开始了交流并意识到这个可能的差错；或者其中一名机组发现了此时的高度与平时给出的高度不一致，立即向 ATC 确认高度指令并纠正了差错。在这种情况下，对于该关键胜任力的评分不应为 1 分，建议参考分为 3 分，也可取决于当天训练的其余部分评为 2 分或 4 分，但可以明确的是这个胜任力不应是 5 分。然而，这并不妨碍其他胜任力有可能根据这个特定场景的元素的证据被评为 5 分。PRO 不应评为 5 分，因为飞行员没有相互确认以及与 ATC 确认，没有落实 SOP 的要求。但是，SAW 有可能为 5 分，因为一名机组突然意识到这个高度与平时的不一致。

（b）无意不合规，产生后果，导致非预期的航空器状态（以下简称 UAS）。

由于机组及时将差错管理成功转化为对 UAS 的管理，有效减轻了差错导致的安全后果。比如，选择错误的高度后，采取由 ATC 指挥避让或按 TCAS 执行避让。对于该关键胜任力的评分不应为 1 分或 5 分。最合理的参考值为 2 分，因为该情况的结果并非不安全（因此不应是

1 分), 仅限于四种可能的等级 (2、3、4 或者 5 分)。这时教员应当考虑: **是否降低了安全裕度?** 如果是, 比如该样例中航空器进入了 UAS, 则评分不能为 5 分。此时此刻, 就在这个过程中, 教员可确定评分既不能是 1 分也不能是 5 分。接着考虑下一步: **安全裕度到底降低了多少?** 通常, 4 分是不太可能的。在这个过程中, 它将取决于情况的背景 (情况很危险吗?)。通常, 进入一个危险的情景, 证明机组有许多行为指标未能很好地表现出来, 因此, 通常不会评为 4 分, 因为 4 分需要有效地展示出“大部分”的行为指标。接下来, 教员会询问机组, “你们对于 UAS 管理得如何”(以及在本场次中的其他情况)。问题导向会集中在这个情况是怎么解决的, 机组人员是否及时调整策略从对差错的管理转化到对 UAS 的管理, 机组是否展示了较好的规避方案或训练标准。如果机组在本次训练的其他场景中表现的相关胜任力能展示较好的规避方案和训练标准, 就可以评为 3 分, 否则这项能力就评为 2 分。

2. 有意不合规的情况, 但机组及时识别并纠正, 结果安全。

例如, 1000ft 不稳定进近后的复飞, 教员可尝试使用以上方法进行实践。对应于该项关键胜任力的参考分是 2 分, 最多评为 3 分。因为从结果看, 这不是一个不安全的状态, 但 800ft 复飞较大的降低了安全裕度。当然, 这也不是一个安全的状态, 因为在手册中明确要求 1000ft 不稳定需要执行复飞, 因此不应评为 4 分或 5 分。另外, 参考本场的其他情况, 存在评为 1 分的可能。但是, 这种方法没有考虑到存在机组为了更高的安全裕度而偏离 SOP 的情况, 如 1200ft 不稳定, 机组为了更好地完成复飞动作, 开始做复飞简令, 最终在 900ft 执行复飞。

如果机组继续不稳定进近并最终落地, 那关键胜任力只能是 1 分, 而且这个关键胜任力很有可能是 PRO, 同时其他的任何一项胜任力都不可能为 5 分 (前提: 没有出现为了提高安全而偏离程序的情况, 比如五边火警、双发失效等)。

对于不合规的情况，要认真分析，每个不合规都是飞后讲评较好的切入点，出现多次无理由的不合规观察项，可以考虑 **PRO** 评为 1 分。

在评分时，**UAS** 常被看作一个结果，其实 **UAS** 是一个介于正常状态与最后状态的中间状态，所以在考虑最后结果时，不应把 **UAS** 当成最终状态。**UAS** 是一个必须管理的中间状态，也是评估的关键所在。

附件 3：EBT 新教员常见评估失准情况分析

新教员很容易将关键胜任力归因到 KNO 和 PRO。比如，在高工作负荷中忘记在云中打开防冰，教员应当将该关键胜任力归因为 WLM，或者机组是手动飞行状态，即可能是 FPM，而新教员很容易将其归类到 PRO。除非机组确实不知道在云中要使用发动机防冰（需要通过其他方式确认），只有这样才可以归类到 PRO。

KNO 的能力确认也类似，所有的能力都是知识技能态度的组合，所以教员在归类到 KNO 时一定要多考虑能力的相关性，即能力金字塔，一定要明确关键胜任力，否则会导致知识归因泛化。

新教员容易仅按照安全结果论去发现问题，往往忽视那些没有产生结果的行为对应的相关胜任力。实际上在较多的 TEM 模型中，飞机和运行系统设计的裕度就能解决一些问题，既然 EBT 训练的目的是提升机组胜任力，就必须关注对于 HOW MUCH 和 HOW OFTEN 两个维度的评估。