

一条路没有交接点

论课堂的达标判定为何由铺路的那一方作出，及一个可从既有档案算出的独立结算比会读到什么

教育学 × 物理学 × 工程学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅 约2.1万字 · 30页

发表 2026年8月3日

摘 要

课堂上真正的学习为什么很少发生？三个行当各给了一套完整而互不相容的做法：工程学把成果结算在造出来的那个东西上，物理学结算在那条能生成一族现象的组织方式上，教育学结算在学习者此后的可能性条件上。三条处方在同一间教室里不能同时执行。本文主张三家都不错，错的是它们共有而未说出的假定：**一段成果可以在某一个位置上单独结算**。推翻它的材料来自工程学自己——它把独立性写进了标准：执行确认的一方必须在技术、管理与经费三方面独立于开发方，三项不全就叫验证与确认、不叫独立的验证与确认。工程学用一份文件承认了，**它的终点不能由走这条路的人来结算**。物理学用盲分析承认了同一件事，其历史依据是同一个量的公布值曾一段一段地朝前一次靠拢；而教育学把这件事的反面写成了核心原则——同一方推导目标、设计活动、并评价与报告表现。于是终点不是位置的属性，是次序上的角色；由此需要一个位置：前一条次序的终点成为后一条的起点、执行方在此更换、判据先于成果存在——**交接点**。**一条路没有交接点，而课堂只有一条路**。文中给出一个可从既有档案算出、不看学习者内部的读数：**独立结算比**，并论证一门同时要求「三者由课程负责人对齐」与「效度须有外部证据」的学科，在不承认这个变量时无法同时守住这两条。

这一篇由哪三个学科的理论体系撞成

第一家主张成果只在造出来的东西上结算——理论正确、过程漂亮、文档齐全都不算数，能不能用是唯一判据；第二家主张成果只在那条能生成一族现象的组织方式上结算——某一次读数与某一台装置都是手段，造不出东西也无妨；第三家主张成果只在学习者此后的可能性条件上结算——当下的表现与那套教法都是手段。三家不能同为真，且各自的日课正是另一家诊断出的病：工程学要把成果拿到教室外面去检，物理学要把判据在看到答案之前锁死并交给不知情的人执行，而教育学的核心原则恰恰要求目标、活动与评价三者由同一方对齐——**同一个动作，在两个行当里是必须被禁止的风险源，在第三个行当里是核心原则**。三家均为站外的公开文献，链接直达原始出处，可自行核对。

工程学 · 验证与确认，及被写进标准的三项独立性

[NASA <i>IV&V Overview</i>](#)（验证问「我们把产品造对了吗」，确认问「我们造对了产品吗」；独立性含技术、管理、经费三项）；标准出处 [IEEE Std 1012, <i>Standard for System, Software, and Hardware Verification and Validation</i>](#)；另见 [Boehm, Verifying and Validating Software Requirements and Design Specifications \(<i>IEEE Software</i> 1:1, 1984\)](#)

主张成果只在「能不能用」上结算；而这一支自己把独立性写进标准——执行确认的一方必须在技术、管理与经费三方面独立于开发方，三项不全就不称为独立的验证与确认。****这正是推翻共有假定的那件材料：它用一份文件承认了自己的终点不能由走这条路的人来结算。****

<https://www.nasa.gov/ivv-overview>

物理学 · 盲分析，与公布值排队朝前一个靠的那段往事

[Klein & Roodman, Blind Analysis in Nuclear and Particle Physics \(*Annual Review of Nuclear and Particle Science* 55, 2005, pp. 141–163\)](#)

主张成果只在那条组织方式上结算；而这一支自己建立了盲分析——判据必须在看到结果之前锁定，其历史依据是同一物理量的历年公布值呈一段一段的台阶、每段内部却互不相容。它防的不是坏人，是结构：**同一方既定判据又读数据，这个组合本身不携带信息。**

<https://doi.org/10.1146/annurev.nucl.55.090704.151521>

教育学 · 建构性对齐

[Biggs, Enhancing Teaching through Constructive Alignment \(*Higher Education* 32:3, 1996, pp. 347–364\); 另见 Trigwell & Prosser, Qualitative Variation in Constructive Alignment in Curriculum Design \(*Higher Education* 67:2, 2014\)](#)

主张成果只在学习者此后的可能性条件上结算。其核心原则要求同一套框架指导各阶段的决定——推导课程目标、决定教学活动、并评价与总结报告学生表现；而这一支自己的迁移研究又表明，迁移很少自发发生、且随情境距离衰减——也就是说它的终点只有在另一个情境里显形时才可验证，同样不能自我结算。

<https://doi.org/10.1007/BF00138871>

目 录

摘 要.....	5
一、导论：同一间教室，三套不能同时执行的做法.....	6
二、三家各自的落点与它们的完整次序.....	7
三、冲突定位：一个具体处境里，三条处方立刻分岔.....	9
四、三家各自知道、却不往这个方向用的那件事.....	9
五、共有假定：成果可以在一个位置上单独结算.....	11
六、六条完整次序、三个终点，与交接点.....	12

七、两根轴与被打中的那一格13

八、右上格：换了执行者而没换判据.....14

九、十二处兑现15

十、与最近的几种说法的分界16

十一、与本栏已有几篇的分界18

十二、剥去学科名词之后18

十三、三家为什么各自到不了这一条.....20

十四、一门不承认这个变量的学科无法同时持有它自己的两条要求20

十五、三家反过来加给本文的约束21

十六、独立结算比：一个可以从既有档案里算出来的读数22

十七、可插手的那一步23

十八、伦理与证据边界24

十九、证伪条件与一条写死日期的赌注.....25

二十、不适用的边界26

二十一、本文自己26

结 论.....27

注 释.....28

摘要

课堂上真正的学习为什么很少发生？工程学、物理学与教育学各给了一套完整而互不相容的做法：工程学把成果结算在造出来的那个东西上，物理学结算在那条能生成一族现象的组织方式上，教育学结算在学习者此后的可能性条件上。三家各自的次序都完整，三家的处方在同一间教室里不能同时执行。本文主张三家都不错，错的是它们共有而未说出的那个假定：**一段成果可以在某一个位置上单独结算，另两个位置只是通往它的手段。**

推翻这个假定的材料来自工程学自己。工程学把"验证"（造对了吗，对照规格）与"确认"（造对了东西吗，对照真实需要）分成两件事，并且把独立性写进了标准：执行确认的一方必须在技术、管理与经费三方面独立于开发方；不具备这三项，那就只是验证与确认，不是独立的验证与确认。**工程学用一份标准承认了：它的终点不能由走这条路的人来结算。**物理学以另一种形式承认了同一件事——盲分析要求分析判据在看到结果之前锁定，其历史依据是数据组的公布值曾长期朝前一次的值靠拢。而教育学把这件事的反面写成了自己的核心原则：建构性对齐要求同一方推导目标、设计活动、并评价与报告表现。

于是本文给出的不是一个更好的终点，是一个位置：**交接点**——前一条次序的终点成为后一条次序的起点、且执行方更换的那个位置。一条路没有交接点；课堂只有一条路。由此得到一个可从既有档案算出、不看学习者内部的读数：**独立结算比**——最近若干次达标判定中，判定方与路径给出方不属于同一结算单元的次數占比。零情态词判据一句：**这门课最近一次"达标"的判定，是由谁按谁给出的判据做出的？**结论：真正的学习稀缺，不是路走错了、走不完或走不到，而是课堂把三次结算收在同一方手里，三次结算因此变成一次自我结算，而自我结算在形式上与真结算完全不可区分。

关键词：交接点；独立结算比；验证与确认；盲分析；建构性对齐；自我结算

Title: A Single Route Has No Handover

Subtitle: On why classroom learning is settled by the same party that laid the route, and what an independent settlement ratio would read

Abstract: Engineering, physics and education each offer a complete and mutually incompatible account of why genuine classroom learning is rare: engineering settles on the artefact that works, physics on the organising account that generates a family of phenomena, education on the learner's subsequent conditions of possibility. All three routes are complete; their prescriptions cannot be executed together in one room. This paper argues that none of them is wrong, and that what is wrong is the premise they share without stating: that an outcome can be settled at one position alone, the other two serving as means. The overturning material comes from engineering itself. IEEE Std 1012 separates verification from validation and writes independence into the standard — technical, managerial and financial independence of the validating party from the developing party. Engineering has thereby

conceded, in a document, that its own endpoint cannot be settled by whoever walked the route. Physics concedes the same through blind analysis. Education has written the converse into its central principle: constructive alignment requires one party to derive objectives, design activities, and assess. The paper therefore names not a better endpoint but a position — the **handover point** — and one reading computable from existing records: the **independent settlement ratio**.

Keywords: handover point; independent settlement ratio; verification and validation; blind analysis; constructive alignment; self-settlement

一、导论：同一间教室，三套不能同时执行的做法

一门大学专业课。教师写了课程目标，按目标设计了教学活动，按目标出了期末试卷，自己阅卷，最后做了一份课程目标达成度分析，各项达成度都在零点七以上。学生评教九十六分。这门课通过了专业认证的材料审查。

三年后，这批学生里的多数人说不清这门课讲过什么，用人单位反映他们在相关任务上并不比没上过这门课的人更强。

没有人失职。每一个环节都按当前公认的最佳做法办的。

这件事已经被三个完全不同的行当各自诊断过，而三套诊断不能同时为真。

第一套来自工程学。 工程学的判据只有一条：**造出来的那个东西，在真实使用条件下能不能用。** 理论正确、过程漂亮、文档齐全都不算数。工程学把这件事做得极细，细到把"验证"与"确认"分成两个词：**验证问的是"我们把东西造对了吗"**（对照规格）；**确认问的是"我们造对了东西吗"**（对照真实需要）^[1]。工程学知道这两者会分开——一个通过全部验证的系统可以完全无法确认，因为规格本身写错了。于是它给出的诊断是：课堂只做了验证，从没做过确认；课程目标就是规格，达成度分析就是对照规格的验证，而"这个规格对不对"从未被问过。处方：把成果放到真实使用条件里去检，能用才算。

第二套来自物理学。 物理学的判据也只有一条，但落点完全不同：**要的是那条能生成一族现象的组织方式。** 某一次实验的具体读数不是成果，某一台装置也不是成果；它们是手段。物理学接受"造不出东西"，只要那条组织方式被逼了出来。于是它给出的诊断是：课堂教的是一堆可以复述、可以套用的结论，而学生没有被逼着把那条组织方式自己逼出来一次；能用不等于懂，会算不等于会推。处方：把学生放到那个必须自己整理出规律的位置上。

第三套来自教育学。 教育学的判据是第三个：**学习者此后的可能性条件有没有被改变。** 一次考试的表现是手段，一套教法也是手段；终点是这个人从此能进入哪些情境、能承接什么。这就是"能力""素养""发展"这些词所指的东西。于是它给出的诊断是：前两家看的都是当下这一次的产物，而教育要的是这一次

之后的那个人；能造出来、能推出来，若三年后什么也剩不下，那就是没学。处方：让目标、活动与评价三者对齐，围绕预期的能力去设计整门课^[2]。

三家不能同时执行。 工程学要把成果拿到教室外面去检，物理学要把学生按在教室里逼出那条组织方式，教育学要把三者收拢对齐——第一条要求判据来自外部条件，第二条要求判据来自内容本身，第三条要求判据由课程设计者统一给出。教师站在那里，三个都对，三个都用不了。

而更要紧的是第四件事，本文正是从它开始的：

> 教育学每天在做的那件事——由同一方推导目标、设计活动、并评价与报告表现——正是工程学写进标准去禁止的那件事，也正是物理学用一整套制度去防的那件事。

工程学的标准写得毫不含糊：执行确认的一方必须在**技术、管理与经费**三个方面独立于开发方；三项不全，那就只是验证与确认，不叫独立的验证与确认^[3]。物理学的做法叫盲分析：分析判据必须在看到结果之前锁定，其历史依据是粒子数据组公布值的一段难堪往事——同一个量的公布值曾在几十年里一段一段地朝前一次的值靠拢，而每一段内部的误差棒都说这些值互不相容^[4]。

而建构性对齐的原文写的是：同一套框架指导全部环节的决定——推导课程目标、决定教学活动、以及评价并总结报告学生表现^[2]。

同一个动作，在两个行当里是必须被禁止的风险源，在第三个行当里是核心原则。 这不是谁抄错了作业，三家都有各自坚实的理由。本文认为这说明三家共同站在一个没说出口的假定上。

本文要立的判断是一句关于**位置**的话，不是一句关于**终点**的话：一次完整的学习不是走一条路，是同一段内容被三条终点不同的次序各走一遍；每一条次序的终点必须由下一条次序的起点来接住，而接住的那一方不能是给出这条路的那一方。这个位置本文称为**交接点**。**一条路没有交接点。课堂只有一条路。**

路线图：第二节给出三家各自的完整次序与它们的落点；第三节把冲突定位到具体处境；第四节取三家各自知道、却不往这个方向用的那件事；第五节说出共有假定并用工程学自己的标准推翻它；第六节给出六条完整次序与交接点；第七、八节给出被打中的那一格与最容易被误判的那一格；第九节做十二处兑现，其中三处反过来削掉本文原本要主张的东西；第十至十三节依次与最近的几种既有说法、与本栏已有的几篇、与上游学科的同形结构、与三家各自划清界线；第十四节论证一门不承认这个变量的学科无法同时持有它自己的两条要求；此后是读数、可插手的那一步、伦理、证伪、边界与自反。

二、三家各自的落点与它们的完整次序

三家在这里必须被写成**各锁一个终点**：每一家都认为成果落在自己那个位置上，另两个位置只是通往它的手段。这不是把它们写极端，它们各自的文本确实是这么写的——一旦写成“三者都重要、要具体情况具体分析”，这一家就不再是一个立场，只是一次弃权。

2.1 工程学：终点是那个能用的东西

落点：成果是**做出来的那个东西在真实条件下的表现**。规格、理论、过程都是手段。

完整次序：从需要（条件）出发 → 定出规格与路线（组织方式）→ 做出一个能用的东西。或者：从既有的做法出发 → 配上条件与资源 → 做出东西。两条次序不同，终点相同。

它自己最硬的一条：验证与确认是两件事，而且会分开。全部验证通过而确认失败，是工程里最贵的一类失败，因为它发生在最晚、代价最高的时候。工程学为此付出的代价是把独立性写进标准，并且在安全关键系统里花掉项目预算里可观的一块。

它给另两家的诊断：物理学那条“逼出组织方式”的路可以走到底而造不出任何能用的东西——这在工程里叫做一个正确而无用的结论。教育学那条“改变此后的条件”的路根本无法验收——它的终点在未来，而未来不进入任何一次交付。

2.2 物理学：终点是那条组织方式

落点：成果是**那条能生成一族现象的组织方式**。某一次读数、某一台装置都是手段。

完整次序：从看到的现象出发 → 改变条件、做出对照 → 逼出那条组织方式。或者：从一个给定的条件设置出发 → 看现象 → 归纳出组织方式。终点相同。

它自己最硬的一条：读数会被期待污染，而污染是无意识的、发生在尽责的人身上的。物理学为此付出的代价是盲分析：在分析完成之前不许知道答案。它明白这不是道德问题，是结构问题——**同一方既定判据又读数据，这个组合本身不可信，与这一方是否诚实无关。**

它给另两家的诊断：工程学那条路造出来的东西可以完全不被理解——它能用，而没有人说得为什么能用，于是它一旦失效就无从修。教育学那条路的终点无法被检验，因为“此后的可能性条件”不是一个可以直接观测的量。

2.3 教育学：终点是此后的条件

落点：成果是**学习者此后的可能性条件被改变**。当下的表现是手段，那套教法也是手段。

完整次序：从当下做出来的样子出发 → 反推出组织方式 → 使此后的条件被改变。或者：从一套教法出发 → 做出结果 → 使条件被改变。终点相同。

它自己最硬的一条：目标、活动与评价不对齐时，学生会学错东西——他们学的是评价实际奖赏的那一样，不是大纲写的那一样。这一条有大量证据，而且是教育学过去三十年最有实践力的一条。建构性对齐正是为解决它而提出的。

它给另两家的诊断：工程学的“能用”是一次性的——这个人这一次做出来了，不说明他此后能做；物理学的“懂了”同样是一次性的——他这一次推出来了，换个场合他推不出来。两家都把手段当成了终点。

三、冲突定位：一个具体处境里，三条处方立刻分岔

三家两两不能并存。

工程学与物理学：前者说造不出来的理解不算数，后者说造出来而不理解的东西是黑箱。若工程学对，那么"逼出组织方式"是一项奢侈；若物理学对，那么"能用"随时可能在下一个条件里失效，而你不会知道为什么。

工程学与教育学：前者说必须在真实条件下验收，后者说真实条件下的一次成功不构成能力。若工程学对，教育学的终点永远无法结算；若教育学对，工程学结算的是一次表现而不是一个人。

物理学与教育学：前者说判据必须在读数之前锁定、且最好由不掌握答案的人执行；后者说判据必须由课程设计与目标、活动一起对齐。**两家对"判据该由谁给、由谁执行"给出了互不相容的归属——这一对最硬，因为它不是侧重不同，是同一个位置上的相反规定。**

放到一个具体处境上，三条处方立刻分岔。

设一门"数据结构"课，教师要判断学生是否掌握了哈希表。

- **工程学**：给一个真实的任务——把一份三百万行的日志按某个字段去重，在给定内存里跑完。**能跑完就算，跑不完就不算**，用什么方法不管。
- **物理学**：**判据先写死并锁上**——在看到任何一份作业之前，把"什么算掌握"写下来、封存，最好交给一个不参与教学的人保管；阅卷时按那份封存的判据打分，且阅卷者不知道哪份是谁的。至于任务真不真实，不重要。
- **教育学**：**先对齐**——课程目标写"能针对场景选择合适的散列策略并说明取舍"，教学活动就要有让学生做取舍并说明的环节，评价就要考取舍与说明，三者由课程负责人统一设计。

三条不能同时执行。工程学那条要求判据来自课堂之外（真实任务），教育学那条要求判据来自课程内部（对齐目标），两者直接对立；物理学那条要求判据的执行者与给出者尽量分开，而教育学那条的核心恰恰是三者统一在一人手里。

教师站在那里，三个都对。

本文认为，三条之所以不能同时执行，是因为三家共同回答的是"成果该在哪个位置上结算"，而这个问题本身预设了一件三家都没说出口的事。

四、三家各自知道、却不往这个方向用的那件事

要推翻三家共有的假定，材料不能从外面搬。从外面搬一个理由来反对它们，得到的只是第四家的立场——那是加入争论，不是取消争论。材料必须是三家之一**自己已经掌握、已经写进标准或制度、只是没往这个方向用的事实**。三家各有一件，而且形状相同。

4.1 工程学自己知道的那一件：独立性被写进了标准

这是本文最要紧的一件材料，因此写得细一些。

工程学早就知道验证与确认是两件事。真正值得注意的不是这个区分，而是**它对确认这件事的执行者提出了什么要求**。

行业标准把独立性拆成三项，缺一项就不算：

- **技术独立**：执行分析的人不参与被分析系统的开发。它的理由写得很直白——不同的视角能发现那些被开发者的注意力结构性遮住的错误。
- **管理独立**：执行分析的组织必须与开发和项目管理不属于同一个组织，并且**自己决定分析哪一段、用什么方法、按什么时间表、对哪些问题提出意见**。
- **经费独立**：分析工作的预算不由开发方掌握。

标准里还有一句更硬的话：**不具备这三项，那些活动就只是验证与确认，不叫独立的验证与确认**——名称的区别不是修辞，它标出的正是**"这份结论携不携带信息"**。

工程学把这一整套读作**质量保证与风险控制**：为了尽早发现偏差、为了降低事故概率、为了让管理层拿到独立的判断。这个读法当然对。但它还有一种读法，而工程学从不采用：

＞ **工程学用一份标准承认了：它自己的终点不能由走这条路的人来结算**。那个"能不能用"的判定，若由造这个东西的人做出，它在形式上仍然是一次判定，而它不携带任何关于"这个东西能不能用"的新信息。

这句话把共有假定掀了。若终点必须由另一个位置来结算，那么"终点"就不是某个位置的属性——它是一个**次序上的角色**：谁做终点，取决于这一轮从哪儿起、又交给谁。

4.2 物理学自己知道的那一件：公布值曾经排着队朝前一个靠

物理学的那件事更难堪，也更有说服力，因为它是自曝家丑。

同一个物理量，几十年里被不同的实验组反复测量。把历年公布值按时间画出来，本该看到的是一团围绕真值的随机散布。实际看到的是一段一段的**台阶**：某一段时间里的公布值彼此挨得很近，与前一段明显不同，而每一段内部各次测量给出的误差棒都说这些值互不相容。最自然的解释是：**后测的人知道前测的结果，于是在决定"分析到此为止"的那一刻，不自觉地停在了离前值更近的地方**。

物理学为此建立了盲分析：把答案藏起来（藏信号区、加一个未知偏移、按比例抽掉一部分数据），直到全部分析判据锁定为止；有些做法还要求解盲的密钥不在做分析的人手里。

物理学把这一整套读作**防止实验者偏倚的技术**。同样有另一种读法，而物理学不采用：

＞ **它承认了同一件事——由同一方既定判据又读数据，这个组合本身不携带信息，与这一方是否诚实无关**。盲分析防的不是坏人，是结构。

4.3 教育学自己知道的那一件：迁移很少自发发生

教育学的那件事已经被研究了一百年：迁移很少自发发生，而且教得越像目标情境，迁移越好；离得越远，越差。

教育学把它读作一条设计要求："要设计迁移""要做变式""要多情境教学"。这个读法当然也对。但它同样有另一种读法：

> "此后的可能性条件被改变"这个终点，只有在它已经变成另一个情境里的一次具体表现时，才是可验证的。也就是说，教育学的终点必须由另一个位置来结算，而那个位置在它自己的账本上不出现——课程的评价发生在课程内部，"另一个情境"要么不存在，要么由同一位教师设计。

三件材料指向同一件事：三家各自都在自己的领域里承认过"我这个终点不能由我自己结算"，而三家都把它记为一种**技术措施**（质量保证／防偏倚／设计迁移），没有一家把它记为**终点这个角色本身是可以交换的**。

五、共有假定：成果可以在一个位置上单独结算

现在可以把三家的争论写成同一个句式：

> 三家争的是：一段成果**该在哪个位置上结算**——在做出来的那个东西上（工程学）、在那条组织方式上（物理学）、还是在此后的可能性条件上（教育学）。

争的是位置的归属。而它们共同假定了一件事：

> **存在一个位置，成果可以在它上面被单独结算；另两个位置只是通往它的手段。**

把这句话念给三家听，三家都会说"这还用说吗"。它不是任何一家的立场，它是三家的立场得以互相反驳的那块共同地面——只有假定存在一个可单独结算的位置，"是哪一个"才成为一个可争的问题。

而第四节的三件材料把这块地掀了。工程学用一份标准说：这个位置上的结算，必须由不在这条路上的另一方来做。物理学用一整套制度说了同一句话。教育学用一百年的迁移研究说了同一句话，只是没这么写。

于是：

> **终点不是某个位置的属性，是一个次序上的角色。谁做终点，取决于这一轮从哪儿起、交给谁。**

这正是"终点与手段可互换"。而一旦终点可互换，"该在哪个位置上结算"这个问题就不再有答案——不是因为三家都错，是因为**这个问题问的是一个不存在的东西**。

由此，本文的判断可以写出来：

> 课堂上真正的学习之所以很少发生，不是因为路走错了（造出来的东西不合真实需要），不是因为路没走完（那条组织方式没被逼出来），也不是因为路的终点没到（此后的条件没被改变）；而是因为课堂只有一条路，而一条路没有交接点。学习要求同一段内容被三条终点不同的次序各走一遍，每一条的终点必须由下一条的起点来接住，且接住的那一方不能是给出这条路的那一方。课堂把这三次结算全部收在同一方手里，于是三次结算变成一次自我结算——而自我结算在形式上与真结算完全不可区分。

这条判断不能从任何一家单独推出，理由在第十三节。

六、六条完整次序、三个终点，与交接点

6.1 六条完整次序

把三个位置——**做出来的样子、组织的方式、所处的条件**——排成"从哪儿起、经过什么、到哪儿"的完整次序，一共只有六条：

次序	读作	谁在走
条件 → 方式 → 样子	有了这些条件，找出办法，做出东西	工程实作
方式 → 条件 → 样子	拿着这套办法，配上资源，做出东西	工艺生产
样子 → 条件 → 方式	看到现象，改变条件做对照，逼出规律	实验科学
条件 → 样子 → 方式	给定一套条件，观察发生什么，归纳出规律	观察科学
样子 → 方式 → 条件	从做出来的东西反推办法，从此有了新的余地	复盘、师承
方式 → 样子 → 条件	教一套办法，练出结果，从此"学会了"	课堂默认

六条按终点分成三组，各两条：终点是"样子"的两条（工程学的落点）、终点是"方式"的两条（物理学的落点）、终点是"条件"的两条（教育学的落点）。

6.2 课堂只走了其中一条

课堂几乎只走最后一条：**给方法 → 练出结果 → 宣称条件改变了**。

这条路本身没有问题。它有效率，它对新手是必需的，它在三家里也有一家（教育学）承认它的终点是对的。问题在于它是**唯一的一条**，而且——这一点是关键——**它的三步全部由同一方安排**：方法由教师给，练什么由教师定，"条件改变了"由教师宣称。

而另外五条次序在课堂上要么不出现，要么出现时被压成这一条的一个环节。"实验课"名义上走的是第三条（看现象、改条件、逼出规律），实际给了完整的实验手册，于是它退回第六条：方法在手册里，结果在报告里，条件改变由成绩宣称。"项目式学习"名义上走的是第一条，实际的项目、资源与验收标准全部由教师给出，于是同样退回第六条。

不是这些形式假，是它们缺一个东西。

6.3 交接点

两条终点不同的次序之间，前一条的终点必须成为后一条的起点，且执行方在这个位置上更换——本文把这个位置称为**交接点**。

交接点有三个条件，缺一个就不成立：

1. **位置更换**：前一条的终点与后一条的起点落在同一处（例如：工程实作做出了一个东西，实验科学正是从"看到的现象"起步）。2. **执行方更换**：接住的一方不是给出前一条路的那一方。3. **判据先行**：接住的一方所用的判据，在它看到那个成果之前就已经存在，且不由给出前一条路的那一方给出。

三条合起来说的是一件事：**一次结算要携带信息，它必须来自另一条次序的起点。**

而这里有一个结构性的后果，本文认为它是最要紧的一句：

＞ **交接点是一个只在"多于一条次序"时才存在的位置。一条路没有交接点——不是它的交接点做得不好，是在单条路径的组织方式里，压根没有这样一个位置。**

这解释了为什么在课堂里"加强评价""引入第三方""做过程性评价"这一类改动往往无效：它们都是在同一条路上加环节，而交接点不是环节，是**两条路之间**的东西。你不能在一条路上加出一个交接点，正如你不能在一根绳子上打出一个接头。

工程学的独立验证与确认、物理学的盲分析，在这个读法下不是"额外的检查工序"，而是**这两个行当各自建成的一个交接点**：工程学把"做出来的东西"交给"真实使用条件"这条次序的起点；物理学把"读到的数据"交给"事先锁定的判据"——那正是另一条次序的起点。

教育学没有建成任何一个交接点，而且它的核心原则要求不建。 建构性对齐的全部力量来自三者由同一方对齐；一旦引入交接点，对齐就断了。这不是教育学疏忽，第十三节会说明它为什么必然如此。

七、两根轴与被打中的那一格

把一次"达标"判定拆成两件事，各作一根轴。

轴一：判据由谁给出——是授课方，还是非授课方。 **轴二：判定由谁执行——是授课方，还是非授课方。**

	判定由授课方执行	判定由非授课方执行
判据由授课方给出	靶格。 建构性对齐的标准形态：目标、活动、评价三者由课程负责人对齐，自己出题、自己阅卷、自己做达成度分析。三次结算全在一方手里，形式上完美自治。	假独立。 自动判题、助教阅卷、同行监考、校内督导听课。 执行者换了，判据没换—— 最容易被当成独立，而按第六节三条件，它的独立性为零。

判定由授课方执行

判定由非授课方执行

判据由 半独立。教师按统考大纲或行业标准自测。判据

非授课 在外部，读数仍由自己给出；比靶格好，仍不满 **交接成立**。只有这一格。

方给出 足条件二。

靶格是左上格。 一个格值不值得写一篇文章去打，有三条签名，缺一条就说明轴选错了——那样的格子不需要任何框架，谁都看得见。

签名一：短期指标全面改善。 靶格里的课程，其对齐度、达成度、通过率、评教分、教学设计评分全都好，而且是真的好，不是装的。教师确实按目标设计了活动，学生确实按活动被评价了。

签名二：失效不产生任何信号。 没有任何一份教学档案会记下"这次达标是自我结算的"——因为通行的记录格式里没有"判据来源"与"判定执行方"这两个字段。缺一个字段与这个字段填了零，在数据上完全不可区分。

签名三：自我加固。 教学质量保障体系越正规——写课程大纲、写课程目标、做目标与毕业要求的对应矩阵、做达成度分析、做持续改进闭环——三样东西就越紧地收在同一方手里。一门课越是按当前公认的最佳做法办，它就越牢地待在靶格里。这一条最要紧：靶格不是执行不到位的结果，是执行到位的结果。

八、右上格：换了执行者而没换判据

右上格值得单独一节，因为它是实务里最有用的一条判别，也是最常被误判的一格。

一门课引入了自动判题系统：题目由教师出，判定由机器执行，客观、一致、不受情绪影响、不认识学生。听起来这是独立性的典范。

按第六节的三条件，它的独立性为零。判据仍然是授课方给的，机器只是把这份判据执行得更整齐。它消除的是执行的随机波动，不是结算的自我性。一个只会做教师出的那类题的学生，在自动判题下会得到与人工阅卷同样的高分，而两者对"他此后能不能做"这件事同样什么也没说。

同样落在这一格的还有：助教阅卷（判据是教师给的评分标准）、同行监考（判据是教师出的卷）、校内督导听课（判据是教学规范，仍由教学管理方给出，且与授课方同属一个结算单元）、以及一门课的期末试卷由同教研室另一位老师"背对背命题"——后者最接近真独立，而它仍然共享同一份课程大纲与同一份教研室绩效。

于是本文对"独立"给出的定义不落在人上，落在**结算单元**上：

＞ **两方是否独立，判据是它们的成败是否由同一笔账结算。** 同一个人当然不独立；不同的人而同属一个考核单元、共享同一份指标、由同一位领导评价，同样不独立。这正是工程学标准把独立性拆成技术、管理、经费三项的理由——它防的不是"是不是同一个人"，是"是不是同一笔账"。

这一条有一个直接的实务后果：**校内的一切"第三方"几乎都落在右上格。** 教务处、督导组、同行评议、教学委员会——它们与授课方属于同一笔账。真正落在右下格的东西很少，而且往往不在学校里：执业资格考试、行业认证、雇主的技术面试、开源社区的代码合入、比赛的裁判。

九、十二处兑现

类比是"像"，兑现是"在那个领域里也能据此预测一件具体的事，且这件事可以不成立"。以下十二处各给一个可读的量。

一、航天软件。 预测：在同一机构、同一时期的项目里，做了完整独立验证与确认的项目，其在轨异常中"需求缺陷"（造对了一个错的东西）所占比例低于只做内部验证的项目；而两类项目的"编码缺陷"占比无显著差异。前者是确认能查出而验证查不出的，后者两种都能查。这份数据在航天机构的异常库里已经存在。

二、核电站调试。 预测：由独立第三方按事先封存的准则执行的调试项，其"通过而后续需返工"的比例低于由安装方自查的项；而两者的一次通过率无差异甚至前者更低。**一次通过率更低而返工更少——这个组合正是交接点存在的签名。**

三、临床试验。 预测：设有独立数据监查委员会且预注册了主要终点的试验，其阳性结果比例低于未设的；而其结果在后续独立重复中的存活率更高。第一项是已知的，第二项是本文的增量预测。

四、粒子物理。 预测：同一个物理量在采用盲分析之后，公布值序列的"台阶"结构应当消失——即后测值与前测值的距离不再系统性地小于其名义误差所允许的范围。这是一条可以在既有数据组历年公布值上直接回算的检验，**而且它是本文十二处里最便宜的一条。**

五、财务审计。 预测：审计方由被审计方直接聘任并付费的，其出具非标准意见的比例低于由第三方（监管机构、投资人委员会）聘任并付费的；控制企业规模与行业。经费独立这一项在这里可以单独变异，因此它能把三项独立性拆开检验。

六、同行评审。 预测：作者推荐审稿人的稿件，其接受率高于编辑独立指定审稿人的稿件；而两组在发表后的引用与重复成功率上无差异或前者更低。多数期刊系统已记录审稿人来源。

七、建筑质检。 预测：由施工方委托的检测机构出具的合格率高于由业主或监管方委托的；而两组在交付后三年的缺陷投诉率相反。

八、药物疗效研究。 预测：资助方与执行方同属一个结算单元的研究，其效应量系统性偏大。这一条已有大量元研究支持——**本文的增量在于给出它与前面几条同形**：它不是医学的特殊问题，是自我结算的一般后果。

九、驾驶培训与驾照考试。 这是社会上少数已经建成交接点的教育场景：教练教、考官按国家标准考，判据与执行方都在教学之外。预测：驾照通过率与教练的"学员满意度"之间的相关，显著低于大学课程里通过率与评教之间的相关。这一条是**正向对照**——它预测的不是问题，是交接点存在时应当看到的样子。

十、编程训练营。 预测：自报就业率与经第三方审计的就业率之间的差值，随该机构自报口径的自主程度单调上升；而两者在有统一审计口径的机构间趋同。

十一、工程教育认证。 预测：课程目标达成度分析（由课程负责人做）与毕业生用人单位评价（外部）之间的相关系数接近零。**这一条最值得做，因为两套数据在已认证专业的档案里同时存在，而它们几乎从不被放在同一张表上比较。** 第十四节会说明为什么不比较。

十二、模型评测。 预测：开发方自建评测集上的提升，在独立保留集上的衰减幅度，随"评测集设计方与模型训练方是否同属一个结算单元"而系统性变化；控制模型规模与训练数据量。这一条与教育的情形结构完全相同——**出题方即答题方，分数即失去信息**——而它是当前唯一有大规模公开数据可查的场合。

十二处里有九处指向同一个形状：一个由自己结算的读数，会同时向好，而它与任何外部效标的相关趋于零。另三处不是兑现，是反过来削弱本文的，写在第十五节。

十、与最近的几种说法的分界

每一条都必须能写成"若观察到甲，则本文对而该说法错；若观察到乙，则反之"。写不成这个形式的，等于承认被它覆盖。

建构性对齐 (Biggs, 1996)。 这是最近的邻居，也是靶格本身，因此必须写得最公允。它主张：课程目标、教学活动与评价必须对齐，否则学生会学评价实际奖赏的那一样。这个主张是对的，本文不反对它，**本文反对的是它对"由谁对齐"的沉默。**判定：设两组课程，目标与活动完全相同，一组的评价判据由课程负责人给出并执行，另一组的判据由不授课的一方按同一份目标给出并执行；若两组学生在外部效标（后续课程表现、执业考试、雇主评价）上无差异，则本文错而对齐足够；若后一组显著更高，则"对齐"必须被拆成"对齐什么"与"由谁对齐"两件事。

形成性评价与终结性评价的区分。 这一支把评价按用途分成两类：一类用于改进过程，一类用于最终判定。分离线：这个区分落在**评价的用途**上，本文落在**判据的来源与执行方**上，两者正交——一次形成性评价可以完全自我结算，一次终结性评价可以完全独立。判定：把四种组合各测一次外部效标相关，若相关只随用途变化而不随来源变化，则本文错。

教考分离。 这是本文最容易被误认成的东西。分离线是硬的：教考分离通常只换了**判据的给出方**（统一命题），并常常伴随目标与活动的脱节，于是它落在本文的左下格（半独立）而不是右下格；更要紧的是，教考分离的多数实施把"对齐"一起砍掉了，因而带来应试化。**本文主张的是对齐的判据加独立的判定，不是不对齐。**判定：在教考分离的实施把"判据是否与课程目标对齐"作为协变量，若效应主要由对齐度承载而与判定执行方无关，则本文的第二根轴多余。

委托代理理论 (Jensen & Meckling, 1976)。这是最危险的一位，因为形式极像：代理人自我监督不可信。分离线：委托代理讲的是**激励不相容**——代理人有动机偷懒或粉饰；本文讲的是**结构上不产生信息**——即使教师完全诚实、完全尽责、激励完全对齐，自我结算仍然什么也没说。判定：找一个激励高度对齐的情形（例如教师收入与学员若干年后的实际收入挂钩的职业培训机构），委托代理预测问题基本消失，本文预测自我结算的信息量仍为零，即该机构的内部达标判定与学员实际收入的相关仍接近零。这个检验可以做，而且这类机构真实存在。

盲分析与预注册。它既是本文的材料来源，也是一位方法学上的占位者，必须说清本文不只是把它搬到教育里。分离线：盲分析防的是**偏倚**（同一方会不自觉地朝预期方向调整读数），本文说的是**信息**（自我结算不产生新信息，即使完全无偏）。判定：在一个可以做到完全无偏的场合（例如判据完全形式化、由机器执行、无任何主观空间），盲分析预测问题解决，本文预测问题仍在——因为判据仍由授课方给出。第八节的自动判题正是这一格，它是这条分离线的现成实例。

独立验证与确认。同样既是材料也是占位者。分离线：它是一道**工序**——在开发流程之外加一个独立的检查方；本文说的是**两条次序之间的交接**——前一条的终点即后一条的起点。判定：一个做了完整独立验证与确认、但从头到尾只走一条次序（按规格造、按规格验）的项目，按它自己的标准是健康的，按本文仍缺两条腿：那条“组织方式”从未被逼出来，那个“此后的条件”从未被改变。可读数：该项目组在下一个不同类型的项目上的表现，与未做独立验证与确认的对照组无差异。

三角互证。社会科学方法论里要求用多个方法测同一个对象。分离线：三角互证要的是**收敛**（多个方法指向同一结论则可信），本文要的是**次序**（前一条的终点被后一条的起点接住），两者不同——三个由同一方设计的方法可以完美收敛而全部自我结算。判定：设三种测法由同一方设计并执行，收敛度很高；三角互证判可信，本文判信息量为零，可用外部效标区分。

分权（孟德斯鸠一系）。剥去学科名词之后，本文的形式最像它，因此必须点名。分离线：分权讲的是**防止权力被滥用**——它是规范性的、防恶的；本文讲的是**读数的信息量**——它是认识论的，与善恶无关。判定：设想一个完全善意、完全能干、绝不滥权的单一权力，分权理论预测无问题（因为无滥用可防），本文预测它的自我判定仍不携带信息。这是两条理论在同一情形上给出相反结论的地方，因此它是一条真分离线，不是“我更强调什么”。

目标置换 (Merton, 1940) 与古德哈特一族。剥完之后本文不是“手段吞噬目的”，也不是“度量一旦成为目标就不再是好度量”——本文说的是**度量根本没有产生，因为它是自己发给自己的**。判定：把评价目标从“课程达成度”改成“三年后的迁移表现”而不改判据来源与执行方，古德哈特一族预测问题缓解，本文预测信息量不变。本栏之八已从另一个角度处理过这一族，此处不再展开。

最近的邻居是建构性对齐。本文不能被它吸收，理由是：**它的核心力量恰恰来自三者由同一方对齐，而这正是本文指认的那一格**。一个必须由同一方完成的对齐，无法在它自己的框架内生出一个交接点。

十一、与本栏已有几篇的分界

本栏此前已有五篇处理同一个问题，它们是本文最近的一批参照，必须逐篇分清。**五篇加本篇的共同底色是"课堂上看起来完成了而实际没发生"**，因此分界必须落到可读数上，不能落在措辞上。

《**方程两边都有它**》（之五十四）。 那篇问同一轮之内**那一步该由谁走**（分岔步的归属、首次选择的可逆性）。本篇问**结算该由谁做**。判决：一堂课里所有分岔步都由学生自己走、首次作答可撤回——那篇判健康——而**判据与判定仍全在授课方手里**，本篇判自我结算，成果与外部效标相关仍为零。

《**每一轮都从新鲜表面开始**》（之五十五）。 那篇的病灶在**轮与轮的接缝**（未完成的中间产物被清零，下一轮从起点重来）。本篇的病灶在**次序与次序之间**（前一条的终点没有被另一条的起点接住）。两者的关键差别：那篇的接缝在**时间上**，本篇的交接在**执行方上**。判决：一个每轮都明确接上上一轮未完成处、结转率很高，而**全程由同一位教师给判据并判定的课**，那篇判健康、本篇判仍无交接。

《**先动的总是同一边**》（之五十六）。 那篇讲**验收的时点**（节拍比：验收间隔与更换做法所需时长之比）。本篇讲**验收的执行方与判据来源**。两条读数正交，可同批测量。判决：一个验收间隔足够长、落差由学生自己读到——那篇判可能有学习——而**所有验收仍由授课教师按自己给的判据做**，本篇判自我结算；反之，验收由校外机构按外部标准做而间隔极密，本篇判有交接、那篇判仍无。

《**记为负项的那一类改变**》（之五十七）。 这一篇与本篇最容易混，必须写清。那篇讲**判据能不能被本轮产物改**（溯改：判据的可修订性，时间维）。本篇讲**判据由谁给、判定由谁做**（结构维）。两侧判决都写死：一个判据完全冻结、永不修订，而验收由独立第三方按公开判据执行的课程——那篇判溯改率为零、是病，本篇判独立结算比高、健康；一个判据频繁按学生产物修订、而修订与判定全由同一位教师完成的课程——那篇判健康，本篇判独立结算比为零、仍无真正的学习。两篇对同一份材料给出相反评价，分离线因此成立。

《**换一种说法，它就不见了**》（之五十八）。 那篇问的是"知识学习的本质是什么"（同一份知识在不同使用条件之间的解绑代价）。本篇问的是"在哪一步能插手"。判决：一份知识绑定宽度很高、解绑代价很低——那篇判健康——而这个判断本身若由授课方按自己的变式题做出，本篇判该读数不携带信息。**这里两篇的关系是嵌套的：本篇的交接点是那篇读数的可信性条件**。这是一处诚实的从属关系，不是分离。

一句合起来才说得出的话：五篇分别指出了这一条路上的五处缺失——分岔步被代走、中间产物被清零、慢驱动没被采样、判据不许被改、绑定只有一处；而本篇说的是这五处缺失有一个共同的存在条件：**整条路的结算权在一方手里，因此这五处缺失都不产生信号**。补上交接点不会自动修好那五处，但没有交接点，那五处的任何一处被修好了也无从知道。

十二、剥去学科名词之后

把本文的判断去掉一切本学科名词，只留下"什么对什么做了什么、导致什么"：

> 一个体系同时持有定判据、执行、与判定三种权力，于是它产出的读数不携带任何来自体系之外的信息；而这个读数在形式上与携带信息的读数完全相同，因此体系无法从读数上察觉这一点。

这个纯形式的结构在上游有名字吗？第十节已点了分权、目标置换、委托代理、三角互证四个，逐条给了分离线。还有两个必须点：

控制论里的"必要多样性"（Ashby）。它说：控制者的状态多样性必须不少于被控对象的扰动多样性。形状相似——一个系统不能控制比自己更复杂的东西。分离线：必要多样性讲的是**容量**（不够多样就控制不住），本文讲的是**来源**（不论多复杂，来自自身的读数仍不含外部信息）。判定：给自我结算的一方增加任意多的评价维度与工具（提高其多样性），必要多样性预测改善，本文预测与外部效标的相关不变。**这是一条可以做的检验，而且它正是当前教学质量保障体系正在做的事——不断增加评价维度。**

信息论里"一个系统对自身的描述"。这一支说得更直接：一条消息的信息量取决于接收方的不确定性被消除了多少；一份由发出方按自己的编码判定的消息，对发出方而言不确定性本就为零。分离线：这一支讲的是**信息量的定义**，本文给的是一个**可在制度上清点的读数**，且指出这个读数与既有的教育指标正交。这一位是本文形式上最近的上游，本文相对它的增量只有一件：**把它落成一个可在既有档案里算出来的比率，并给出它与已有指标正交的实例。**

到这里可以说一句更重的话。上面这些概念——从建构性对齐到形成性评价到独立验证与确认到教考分离——它们把什么当作了给定，因而看不见什么？

- 建构性对齐把**对齐是好的**当作给定，因此看不见对齐要求同一方持有三样权力。一旦承认这一点，它的核心原则就与它自己追求的证据质量冲突。
- 形成性评价把**评价是反馈**当作给定，因此看不见评价同时是结算。一旦承认，形成性与终结性的区分就不再是用途不同。
- 独立验证与确认把**独立是一道工序**当作给定，因此看不见它其实是两条次序之间的交接。一旦承认，在流程末尾加一道检查就不够了。
- 教考分离把**独立与对齐是一根轴的两端**当作给定，因此看不见它们是两根轴。一旦承认，历次教考分离改革的失败就有了解释。
- 委托代理把**问题在动机**当作给定，因此看不见结构。一旦承认，设计更好的激励就不再是完整处方。

这几行背后是同样被当作给定的东西：

> 一个体系可以自己判断自己是不是达到了它自己设定的目标，而这个判断携带信息。

这块地基不属于任何一派——主张外部评价的与主张校本评价的都站在它上面，它们只争"由谁判断更准"，从不问"自我判断这件事本身产不产生信息"。**"没有被任何学科辨识过"的真正来源，不是没有人看，是看的人都站在它上面。**

十三、三家为什么各自到不了这一条

三家没能走到这一步，不是因为疏忽。一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子。

工程学到不了，是因为它只有一个终点。 工程学建成了交接点——独立的验证与确认——但它建的是**通向同一个终点的一道保险**：不论谁来验，验的都是"这个东西能不能用"。工程学没有第二条终点不同的次序需要交接，因此它没有动机去问"交接点是不是一类位置"。它把独立性写成了一条工序纪律，而不是一个位置的性质。**它有形式而没有问题。**

物理学到不了，是因为它的对象不会因为被判定而改变。 盲分析防的是研究者，不是被研究者：电子不会因为被测而努力表现，实验数据不会为了通过评审而调整自己。于是物理学从不需要问"这次判定对被判定者做了什么"。而课堂里，被判定的是一个会朝判据生长的人——这正是教育学那条"学生学的是评价实际奖赏的那一样"所说的。**物理学有制度而没有第二个主体。**

教育学到不了，是因为它的终点在未来，而未来不能验收。 教育学的落点是"此后的可能性条件"，这个终点在课程结束时还没有到来。要在课程之内结算它，唯一可行的办法就是**找一个当下的代理量并让它与目标对齐**——这正是建构性对齐。也就是说：**对齐不是教育学的一个选择，是它的终点落在未来这件事逼出来的唯一可操作方案。** 它不是没想到独立结算，是它一旦独立结算就没有任何东西可以在课程之内被结算。

三家的到不了形状不同，结果一致：三家各自都建成或需要过交接点的某一半，而没有一家把它当成一个位置。

十四、一门不承认这个变量的学科无法同时持有它自己的两条要求

这一节要做的事比划界更重：不是说明本文与既有说法不同，而是说明**不承认本文的变量，教育质量保障领域自己已有的两条要求就不能同真。**

这个领域目前同时要求：

甲：课程目标、教学活动与评价必须由课程负责人对齐，并据此做课程目标达成度分析。这写在几乎所有成果导向的专业认证要求里。

乙：评价的效度需要外部证据——分数必须与外部标准（后续表现、执业考试、用人单位评价）相关，否则它测的不是它声称在测的东西。这是测量学的基本要求，同样写在认证要求里。

两条各自都有理由。要它们同真，必须存在一个判据，把"由对齐带来的一致性"与"由效度带来的一致性"分开。

而这两者在自我结算下**在形式上完全不可区分**：一门完美对齐的课程，其评价与目标高度一致——而这个一致性恰恰是设计出来的，不是发现的。用它去论证效度，等于用设计的结果去证明设计是对的。

于是有一个后果：

> 要让甲以现在的形式成立，必须先承认本文的变量（判据来源与判定执行方是否同属一个结算单元）；而一旦承认，甲就不能再以现在的形式成立——它必须改写为"由课程负责人对齐判据，由不同结算单元的一方执行判定"。

这已经在现行处方里发作。成果导向的认证同时要求两套数据：课程目标达成度分析（由课程负责人做）与毕业要求达成度的外部佐证（用人单位与校友问卷）。两套数据在同一份认证材料里同时存在，而它们几乎从不被放在同一张表上做相关分析。若它们高度相关，达成度分析就得到了外部支持；若接近零，达成度分析的信息量问题就会当场暴露。认证标准对这个相关分析的沉默不是疏忽，是因为做这件事需要承认一个它不承认的变量。

本文在这里给自己留一条退路，写在证伪条件第八条：若课程目标达成度与外部效标的相关系数被实测证明显著为正（例如超过零点四），则本节的主张失败，甲乙两条可以在不承认本文变量的情况下调和。第九节第十一处正是为这条退路准备的检验，而所需数据在已认证专业的档案里已经存在。

十五、三家反过来加给本文的约束

兑现如果只朝一个方向走，说明本文只是在借别的领域打扮自己。以下三处是反过来削掉本文原本要主张的东西的，每一处都写清削掉了什么。

工程学削掉的："所有验收都应当独立"。

这句更强的话本文原本会写下，而工程学不允许。独立的验证与确认很贵——它只在安全关键与任务关键的系统里全面实施，在一般项目里被认为不值当；它还会拖慢进度、在开发方与验证方之间制造对抗关系，而这种对抗本身有代价。

于是本文的处方被削为：**独立结算比有一个由代价决定的上限，它应当被用在"错了以后代价不可逆"的那些位置上，不是全面铺开。**具体到课堂：一门课里真正需要交接点的往往只有一两处（那些"过了这一关此后就默认你会了"的位置），不是每一次作业。这一处动的是适用范围本身。

物理学削掉的："判据必须来自外部"。

本文的第一根轴原本要求判据由非授课方给出。物理学不允许——盲分析里的判据通常正是由做实验的一方自己定的，它的独立性来自**时点**：判据在看到结果之前锁定并公开，此后不许改。

这一处真削弱了本文的第一根轴：**"外部给出"不是唯一的实现方式，"事先锁定并公开"是一个便宜得多的替代。**核心不是"谁给判据"，是"给判据这个动作与读判据这个动作不能是同一次动作"。本文接受这个削

弱，并认为它使处方更可行：一位教师在学期初把"什么算掌握"写死、公开、封存，此后不改，就已经建成了半个交接点，成本近乎为零。

但物理学同时给出了这个替代的代价，本文一并写下：**锁死判据会让人错过真实的异常**——一个事先没想到的、更有价值的东西出现时，锁死的判据读不出它。这正是第十四节所说的溯改与本文的张力，本栏之五十七专门处理了这一面。

教育学削掉的："独立越高越好"。

本文原本会把对齐写成问题，而教育学不允许。对齐解决的是一个真问题：目标、活动与评价不对齐时，学生会学评价实际奖赏的那一样，而不是大纲写的那一样——这有大量证据。完全的独立结算会带来教考分离的老毛病：考的不是教的，学生无所适从，且导向应试。

于是本文的价值排序被削为：**对齐与独立不是一根轴的两端，是两根轴。** 要的是**对齐的判据加独立的判定**，不是不对齐。这一处同时解释了为什么历次教考分离改革常常失败——它们把对齐一起砍掉了，于是用一个问题换了另一个问题。

十六、独立结算比：一个可以从既有档案里算出来的读数

第八节已经把"独立"定义在结算单元上而不是人上。据此可以给出一个读数：

＞ **独立结算比 = 最近 N 次"达标"判定中，判据来源方与判定执行方均不与授课方同属一个结算单元的次数占比。**

分子分母都可以清点。计入的是**进入任何形式的记录与比较的判定**——不按名称计，一次改叫"练习"而仍然记分并被用于比较的检查，仍然算一次判定。

这个读数在三个领域里量纲不同而比值相同：

- **教育**：由非授课方按非授课方给出的判据做出的达标判定，占全部达标判定的比例。
- **工程**：由技术、管理、经费三项独立的一方执行的验收条目，占全部验收条目的比例。
- **物理**：在事先锁定并公开判据的条件下完成的分析条目，占全部分析条目的比例。

三处的单位都不同，比值都无量纲，可以并排比较。

正交性检查。 一个读数若与既有主要指标高度相关，它会被立刻当成那个指标的一个代理，白造。要证明它不是，必须举得出一个"既有指标最好看、而这个数最难看"的真实例子。举得出，而且这个例子就是常态：

＞ **一门对齐度评分最高、达成度全部达标、通过率九成八、评教满分的课，其独立结算比恰为零。** 原因不是巧合——对齐度评分高，正因为目标、活动与评价三者由同一方设计得严丝合缝。**这两个指标不只是正交，方向相反。**

零情态词判据一句：

> 这门课最近一次"达标"的判定，是由谁按谁给出的判据做出的？

这句话可以拿去问教务档案、评分记录、命题记录，不需要问任何人的判断。把它在五个场合各跑一遍，答案互不相同：

1. 大学专业课期末考：授课教师命题、授课教师阅卷、按自己写的课程目标。**零**。 2. 全国统一考试科目：命题方与授课方分离，判据在课程之外。**高，但仍在同一个教育系统内**——命题方与授课方的成败仍由同一套升学率结算，因此不是满分。 3. 驾驶考试：教练教，考官按国家标准考，两方分属不同结算单元。**高**。 4. 博士论文答辩：外部委员评审，判据在学位标准里。**中偏高，而不是很多人以为的高**——委员由导师参与遴选，且同行互评的往复关系使双方常常落在一个更大的结算单元里。 5. 编程训练营的项目：学员做，雇主在技术面试里按自己的用人标准判。**最高**——判据来自另一个条件场。 6. 一门课的自动判题系统：出题方即授课方，机器执行。**零**，而它在所有质量报表上看起来是独立的。

第 4 与第 6 两条不是从命题推出来的，是查出来的——它们都与直觉相反，也都是本文最容易被推翻的地方。

十七、可插手的那一步

这是一道"该怎么办"的问题，因此本文欠一个答案：在这条路上，哪一步可以插手。

答案不在内容，不在难度，不在时间，也不在教法。**在交接点。**

具体是三件事，按代价从低到高：

第一件，最便宜的：把判据先锁上。 学期初把"什么算掌握"写死、公开、存档，此后不改；期末按那份存档的判据判定。这一件成本近乎为零，教师本人仍是判据的给出者——它建成的是半个交接点（第十五节里物理学给出的那个替代）。它的效力上限也因此有限：它挡住的是"事后按学生实际做到的样子调整标准"这一类，挡不住"判据本身从一开始就只指向我教的那一样"。

第二件，中等代价的：把判定交出去一次。 一门课里选一到两处"过了这一关此后就默认你会了"的位置，把判定交给一个不与授课方同属一个结算单元的一方——校外的行业认证、真实项目的接收方、另一所学校的对口教师、开源社区的合入评审。**只交一到两处，不是全部**，这是第十五节里工程学给出的约束。

第三件，最贵也最完整的：让下一条次序的起点来接。 前一条次序的终点（做出来的那个东西），成为下一条次序的起点（看到的现象），由走下一条次序的人接住。落到课堂上，最朴素的形态是：**这一届学生做出来的东西，成为下一届学生要解释、要改进、要推翻的对象，而下一届不是这一届的作者。** 这一件之所以贵，是因为它要求课程在时间上跨过一个学期，而现行的课程单元恰好以学期为边界——**这个边界正是交接点在结构上不存在的直接原因之一。**

三件里，本文只对第一件有把握说它成本可承受、可单独检验；第二、三件都需要制度上的安排，本文不能仅凭推理主张它们值得做。

这三件都不落在教学法上，这是本文与前面五篇的又一处分界：那五篇的干预都还在教师能单独执行的范围内（改归属、改结转、改节拍、改判据的可修订性、改绑定宽度），本文的干预有两件不在。这既是它的力量，也是它最可能落空的地方。

十八、伦理与证据边界

独立结算比是一个比率，而比率几乎必然会被拿去考核人。三条必须写死。

第一，这个读数指的是位置，不是人。它读的是一门课程的组织方式与一个岗位的结算安排，不是这位教师的可信度。**它为零不表示这位教师做了任何不当的事——恰恰相反，在最尽责、最认真做对齐的教师那里，它同样为零，而且往往更低。**把这个数记在个人头上，是把结构问题记成人品问题，而这正是本文第十节对委托代理那条分离线要防的误读。

第二，不得为了提高这个数，而在教师须对学生安全负责的情形里把判定权外包给不在场的一方。实验安全、临床带教、体育训练、野外实习——这些场合的判定必须由在场且负责的人做出，独立结算比在这里应当低，且不应被要求提高。

第三，已按这个读数做出不利安排（评级、经费、认证结论）的，必须公开"同一结算单元"的界定规则并给出复核通道。什么算同一笔账，有相当大的界定空间；不公开界定规则的比率不可用于任何人事决定。

还有一条属于本文自己的反噬预言，必须在正文里写明：

> **这个读数一旦进入考核，最省事的达标办法是成立一个形式上独立、实质上互相评判的委员会——**甲校的老师评乙校，乙校评甲校；或者同一个联盟内部轮流互评。它在编码上完全合规，成本比真独立低一个数量级，而它的结算单元其实是同一个。这件事在学术界有现成的先例，不必等它发生。

我看不到出口。能做的只有一件：把"独立"的定义死死钉在**成败是否由同一笔账结算上**，而不是钉在"是不是同一个人""是不是同一个单位"上；并要求任何声称独立的安排公开它的结算关系。这提高了做手脚的成本，但没有取消它。

十九、证伪条件与一条写死日期的赌注

先写出最强的竞争解释，不挑好打的靶子。最强的一句是：“**其实不就是教师太忙、评价工具做得糙吗？**”——把评价工具做好、投入做够，问题就解决了。次强的两句是“其实不就是激励不对吗”（委托代理）与“其实不就是教考没分离吗”。以下条件必须能排除它们。

第一条（核心）：控制教师投入、评价工具质量（信度、区分度、命题规范性）与激励结构后，若“判据来源方与判定执行方是否与授课方同属一个结算单元”和“课程成绩与外部效标（后续课程表现、执业考试通过率、用人单位评价）的相关”之间不存在剂量-反应关系，则本机制为伪——届时应归因于评价工具质量而非结算结构。样本纪律：**在同一所学校、同一专业、六门以上课程内做**，只在独立结算比上有变异；**严禁拿两个国家或两个地区的对比充数**，那种比较在方法上控制不住任何东西。

第二条（排除委托代理）：在一个教师收益与学员若干年后实际表现挂钩的机构里（这类职业培训机构真实存在），若自我结算的读数与外部效标的相关显著恢复，则激励解释足够，本文为伪。

第三条（两根轴是不是都必要）：把“执行方换而判据方不换”（自动判题、助教阅卷）与“两者都换”分成两组对照，若两组的外部效标相关无差异，则本文的两根轴中有一根多余，第七节那张表应当塌成一根轴。

第四条（顺序，不是相关）：交接点的建立应当早于外部效标相关的提升出现，且间隔不小于一个课程周期。若相关的提升先于或同时于交接点的建立出现，则本文的次序判断为伪。

第五条（最便宜的一条）：在已通过成果导向认证的专业里，“课程目标达成度”与“用人单位评价”两套数据同时存在于既有档案中，把它们做相关即可。所需数据不需要任何新的采集。若相关系数显著为正（超过零点四），第十四节的强制不一致失败。

第六条（正交性）：若独立结算比与通过率或评教分的相关系数超过零点五，则它只是既有指标的一个代理，白造。

第七条（物理侧，可回算）：若在采用盲分析之后，同一物理量公布值序列的台阶结构没有减弱，则“自我结算导致读数系统性靠拢”这一机制在物理侧不成立，本文的材料基础被削弱一半。**这一条的数据全部公开，可以立刻做。**

第八条：若第十四节的甲乙两条能在不承认本文变量的情况下被调和（给出一个把设计的一致性与效度的一致性分开的其他判据），则强制不一致失败。

第九条：若在一个判据完全形式化、执行完全无主观空间的场合（自动判题）里，外部效标相关与人工阅卷无差异且都很高，则本文关于“自动判题落在假独立格”的判断为伪。

本文可以被分层引用。第一层是“终点是次序上的角色、交接点是一类位置”，最强也最易倒；第二层是那张两行两列的表，作为分析工具；第三层是一条不依赖前两层的经验主张——**自我结算的达标读数与外部效标的相关接近零**。第三层最稳也最便宜，第五条证伪条件正是为它写的。第一层被推翻时，后两层仍然站得住。

赌注。 到二〇三一年十二月三十一日为止，至少三项独立研究报告：在教师投入、评价工具质量与激励结构匹配的条件下，仅改变"判据来源方与判定执行方是否与授课方同属一个结算单元"，即产生课程成绩与外部效标相关的差异，方向为分离组更高。

什么不算命中，一并写死：仅实行教考分离而不问判据是否与课程目标对齐，不算；仅更换阅卷人或引入自动判题，不算；仅由校内督导、教务处或同教研室教师执行判定，不算（它们与授课方同属一笔账）；仅报告通过率、评教分或满意度，不算；机构自称已引入"第三方评价"而不公开结算关系，不算；与其他改革打包实施而无法分离出本变量，不算；仅在实验室情境成立而无真实课程重复，不算。

若本文被证伪，学到的是什么？ 若第一条失败而第五条成立——即结算结构无关，而达成度与外部效标确实相关——那么现行的质量保障体系是有效的，问题应当回到教学法本身，本栏前面五篇的方向是对的而本篇是错的。**证伪的方向指向一个不同的干预，这是它值得写下的理由。**

二十、不适用的边界

第一，安全责任在场的情形不适用。 第十八节第二条已写明：判定必须由在场且负责的人做出的场合，独立结算比应当低。

第二，终点确实只有一个的领域不适用。 有些学习的目标就是"能不能做出规定的那一样"，且这一样有唯一的外部标准（操作规程、急救流程、飞行检查单）。这些领域的交接点由标准本身提供，不需要三条次序。

第三，完全的新手不适用。 在学习者还没有任何可被交接的产物之前，唯一可走的就是那条默认次序。本文说的不是那条路错了，是它不能是唯一的一条。

第四，本文不主张三条次序的先后。 哪一条先走、走多久、在哪里交接，本文给不出根据。这是本文交出去的第一个洞。

第五，本文说不出第一个交接点由谁建立。 在一个从来没有交接点的体系里，提议建立交接点的人本身在体系内部，他的提议同样由这个体系自己结算。这是本文交出去的第二个洞，而且是更深的一个——第二十一节会说它同样落在本文自己头上。

二十一、本文自己

用本文自己的判据看本文，结论不好听，而且比前面几篇更难堪。

本文既给出了判据（交接点的三条件、独立结算比），又用这个判据判定了工程学、物理学与教育学三家的状况，还自己评价了这次判定的价值。**定判据、执行、判定，三样全在一方手里。**按第七节那张表，本文落在左上格，而且落得很深。

本文没有任何机制让它自己的结论被独立结算：它的读者与它属于同一个话语的结算单元，同意它的人会引用它，不同意的人不会去做那个检验。**它只能靠被使用来检验，不能靠被讨论来改进。**

还有一处更难堪的：**本文的处方一旦被采纳，最先被要求提供独立结算的会是课堂，而不是提出这条主张的这类文章。**我把这一条写下来，不是谦辞——它是本文自己判据的一次应用，而且它给出了一个可查的预测：**若本文这类主张此后确实被引入某个质量保障体系，可以去看那个体系有没有对它自己的评价框架也做一次独立结算。按本文预测，不会有。**

我能做的只有两件。第一，把证伪条件写得比结论显眼，尤其是第五条与第七条——它们都不需要任何新的数据采集，谁都可以立刻去做，而它们能推翻本文。第二，在这里写下本文自己的判据来源与判定方式：**本文的判据是我写的，本文对三家的判定也是我做的，两者之间没有任何交接。**读者要判断这篇文章是不是自我结算的产物，只需要问它自己那句判据。

交出去的第三个洞：**本文说不清"同一个结算单元"的边界在哪里终止。**甲校评乙校，若两校在同一个排名体系里竞争，它们是不是同一笔账？追下去，任何两方在足够大的尺度上都属于同一笔账。本文没有给出这个边界的确定办法，只能说它是程度问题而不是有无问题——**而一个程度问题被写成一个比率，恰恰是本文第十八节所警告的那种事。**

结论

课堂上真正的学习为什么很少发生？三个行当给了三套完整而互不相容的答案：工程学说成果要落在能用的东西上，物理学说要落在那条组织方式上，教育学说要落在此后的可能性条件上。三条处方在同一间教室里不能同时执行。

本文主张三家都不错。错的是三家共有而未说出的那个假定：一段成果可以在某一个位置上单独结算，另两个位置只是手段。

推翻它的材料来自工程学自己：它把独立性写进了标准——执行确认的一方必须在技术、管理与经费三方面独立于开发方，三项不全就不叫独立的验证与确认。**工程学用一份文件承认了它的终点不能由走这条路的人来结算。**物理学用盲分析承认了同一件事，其历史依据是公布值曾经排着队朝前一个靠。教育学则把这件事的反面写成了核心原则：同一方推导目标、设计活动、并评价与报告表现。

于是终点不是某个位置的属性，是一个次序上的角色。而这这就要求一个位置：前一条次序的终点成为后一条次序的起点、执行方在此更换、判据先于成果存在——**交接点。一条路没有交接点。课堂只有一条路。**

由此得到一个可从既有档案里算出、不看学习者内部的读数：独立结算比。它与既有指标不只是正交，方向相反——对齐度评分最高、达成度全部达标、评教满分的那门课，它恰为零。判据只有一句：**这门课最近一次"达标"的判定，是由谁按谁给出的判据做出的？**

三处约束由三家反过来给出，本文照单接受：独立结算有代价，只该用在错了以后不可逆的位置上；判据不必来自外部，事先锁定并公开是便宜得多的替代；对齐与独立是两根轴而不是一根轴的两端——要的是对齐的判据加独立的判定，不是不对齐。

最后一句留给那件最难办的事。左上格之所以能长期存在，不是因为有人做错了什么，而是因为**它的每一项指标都在改善，它的失效不产生任何信号，而且越是按当前公认的最佳做法去办，它就越牢固。**一个不发信号的失效，只能靠一个原来不存在的字段被发现——这个字段是：**这一次达标的判定，判据是谁给的，判定是谁做的。**

注释

[^1]: 验证与确认的这对提法在软件与系统工程里已成惯用表述：验证问"我们把产品造对了吗"，确认问"我们造对了产品吗"。参见 NASA IV&V 项目的公开说明页对二者的界定，以及 Boehm 对需求与设计规格的验证与确认的早期讨论（IEEE Software, 1(1), 1984）。

[^2]: Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32(3), 347–364. 该文摘要明确写道，建构性对齐是用建构主义作为框架，去指导教学设计各个阶段的决定——推导课程目标、决定被判断为能引出这些表现的教学活动、以及评价并总结报告学生表现。**"各个阶段"这三件事由同一套框架、同一位设计者作出，正是本文第七节所指认的那一格。** 本文不主张该文错，只主张它对"由谁作出"保持了沉默。

[^3]: 独立性的三项参数（技术独立、管理独立、经费独立）见 IEEE Std 1012 及 NASA 软件工程手册相应条目。其中管理独立还要求执行方自行决定分析哪些部分、采用何种方法、按什么时间表进行；经费独立要求预算不由开发方掌握。标准中另有一句本文反复引用的表述：不具备这三项独立性的活动只是验证与确认，不称为独立的验证与确认。

[^4]: Klein, J. R., & Roodman, A. (2005). Blind analysis in nuclear and particle physics. Annual Review of Nuclear and Particle Science, 55, 141–163. 该文给出了若干物理量历年公布值的时间序列图，其形态是一段一段的台阶：某一时期内的各次测量彼此接近而与前一时期明显不同，且被认为彼此影响。作者据此论证实验者偏倚的存在，并系统介绍了隐藏信号区、隐藏答案、增删事件、数据预缩放等盲分析方法。

[^5]: 迁移研究的经典结论（迁移很少自发发生，且随情境距离衰减）见 Detterman & Sternberg 编《Transfer on Trial》(Ablex, 1993) 与 Barnett & Ceci 的迁移分类框架 (Psychological Bulletin,

128(4), 2002)。本文只使用其中"终点须在另一情境中显形才可验证"这一条, 不涉及关于迁移是否可教的争论。

参考文献

- Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128(4), 612–637.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364.
- Boehm, B. W. (1984). Verifying and validating software requirements and design specifications. *IEEE Software*, 1(1), 75–88.
- Detterman, D. K., & Sternberg, R. J. (Eds.). (1993). *Transfer on Trial: Intelligence, Cognition, and Instruction*. Ablex.
- IEEE. IEEE Standard for System, Software, and Hardware Verification and Validation (IEEE Std 1012).
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
- Klein, J. R., & Roodman, A. (2005). Blind analysis in nuclear and particle physics. *Annual Review of Nuclear and Particle Science*, 55, 141–163.
- Merton, R. K. (1940). Bureaucratic structure and personality. *Social Forces*, 18(4), 560–568.
- NASA. IV&V Overview 与 NASA Software Engineering Handbook, SWE-141: Software Independent Verification and Validation.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *PNAS*, 115(11), 2600–2606.
- Trigwell, K., & Prosser, M. (2014). Qualitative variation in constructive alignment in curriculum design. *Higher Education*, 67(2), 141–154.

字数与版本: 约二万汉字, 第一版, 二〇二六年八月。

人机分工声明: 本文的选题、三家的确定与全部判断由王德生提出并定稿; 文献核对、结构组织与文字撰写由 Claude 完成。文中十二处兑现均为可检验的预测, 未经实证, 标注为预测而非结论; 本文不含任何已完成的实验数据。

本文由三个分属不同学科、且互相冲突的理论体系碰撞而成。全部来源均为站外公开文献，可自行核对。 作者：王德生 + Claude · SDE Universes · 学科通融