

学科通融 · 之五十七

记为负项的那一类改变

论课堂中溯改通道的规范性关闭——发起权为何不换手，以及可问责与学习为何要求相反方向的驱动

教育学 × 化学 × 管理学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅 约2.2万字 · 30页

发表 2026年8月3日

摘 要

课堂上真正的学习为何很少发生？三种现行回答互不相容：教学路径不对、认知条件不允许、局部改良被场域吸收。本文主张三者都不是主因。三家争的是那条驱动是哪一条，而三家共同假定了一条谁也没说出口的东西：**谁驱动谁是给定的，在过程里不变**。推翻它的材料来自三家之一自己——一种抗病毒药物更稳定的晶型一旦在某个车间偶然成核，晶种即改写了全球范围的生产条件，原来那条工艺此后再也走不通；化学把它记为污染事故，而它在形式上正是**产物反过来改写了驱动它的条件**。由此命名一类被三家账本一致记在负栏里的事件——**溯改**：由本轮实际产物触发、记名、且约束下一轮的判据修改。在教育里它叫标准不一致，在化学里叫批次污染，在管理里叫标准漂移；三家都有字段、都有处置程序，唯独符号是负的。一类事件被系统地记为负值，等于它作为一样正的东西不存在。**没有溯改，起始端就永远停在同一处；发起权不换手，体系就在第一次选定的那一相里越长越大，而所有指标都在改善**。文中给出可事后清点的比率读数溯改率、一张两轴四格的辨别格，并以中国某试点城市二〇一五至二〇二〇年那场「两次考试取高分」的真实改革作检验——它把首次作答变成了可撤回，四年追踪显示考试频次、教师中心讲授与机械记忆全部上升，项目式活动被放弃：**可撤回性若不伴随溯改，只增加轮次，不改变状态**。最后一句本文不打算缓和：可问责要求规则先于且独立于产物，学习要求产物改写规则，而一间教室一次只能是其中一个。

这一篇由哪三个学科的理论体系撞成

第一家主张决定学生怎么学的是评价制度，教学法只是因变量，故处方是改评价；第二家主张走通哪条路由成核与生长的动力学以及当下条件单独决定，热力学上最稳的那一相不是决定项，故改难度与时长根本不进入路径选择；第三家主张场域标准是被大量组织日复一日的实践改写出来的，任何局部改良都要先被写成一份可申报的材料才能活下去，因而必然被吸收。三家不能同为真，且各自的日课正是另一家诊断出的病：改评价是第一家的日课，而第三家把用评价驱动教学直接列为脱耦的来路；讲究起步条件是第二家的日课，而第一家会说那只是把难度调低的另一种说法；承认场域改不动是第三家的克制，而第二家会说那是把一个动力学问题当成了热力学问题。三家均为站外的公开文献，链接直达原始出处，可自行核对。

教育学 · 反拨效应与它四年之后的走向

[Alderson & Wall, Does Washback Exist? \(<i>Applied Linguistics</i> 14:2, 1993\);](#)
[Huang & Zeng, Sustainability of Washback Effects: A Longitudinal Study of China's Twice-Yearly NMET Reform \(<i>Language Testing in Asia</i> 15:19, 2025\);](#) 另见 [Bennett, Formative Assessment: A Critical Review \(<i>Assessment in Education</i> 18:1, 2011\)](#)

主张决定学生怎么学的是评价制度，教学法是因变量；而这一支自己确立的两件事——反拨对内容的影响远强于对方法，以及课堂活动差异并不由考试设计解释而由教师专业性与信念解释——等于承认它主张的那条驱动在传到方法层时会被别的东西改写。

<https://doi.org/10.1186/s40468-025-00358-9>

化学 · 阶段律的再检验、动力学捕获与消失的晶型

[Cardew, Ostwald Rule of Stages—Myth or Reality? \(*Crystal Growth & Design* 23:6, 2023\)](#); 对立面见 [Threlfall, Structural and Thermodynamic Explanations of Ostwald's Rule \(*Org. Process Res. Dev.* 7:6, 2003\)](#); 案例见 [Bučar, Lancaster & Bernstein, Disappearing Polymorphs Revisited \(*Angew. Chem. Int. Ed.* 54:24, 2015\)](#)

主张结晶出来的是哪一相由成核与生长动力学及当下条件单独决定，热力学稳定性不是决定项；而这一支自己反复综述的一类现象——产物一旦大量存在便通过晶种改写下一批的条件、使原路径全局不可达——正是推翻共有前提的那件材料，只是它把这件事记在污染与工艺事故名下。

<https://doi.org/10.1021/acs.cgd.2c00141>

管理学 · 实践驱动的制度变迁，与它同年的对手

[Smets, Morris & Greenwood, From Practice to Field: A Multilevel Model of Practice-Driven Institutional Change \(*Academy of Management Journal* 55:4, 2012\)](#); 因果方向相反的一篇见 [Bromley & Powell, From Smoke and Mirrors to Walking the Talk \(*Academy of Management Annals* 6:1, 2012\)](#)

主张场域层的逻辑变迁起源于个体在日常工作中的即兴处置，先被证成、再扩散、最后改写场域；而这一支自己写明，这条链条必须经过一道用场域的语言写成的正当化——即它主张的那条驱动，其成立条件由它所驱动的那样东西提供。

<https://doi.org/10.5465/amj.2010.0013>

目 录

论课堂中溯改通道的规范性关闭——发起权为何不换手，以及可问责与学习为何要求相反方向的驱动5

摘要.....	5
一、导论：三个不该同时为真的判断.....	6
二、三家的最新立场，以及各自留下的那一处松动.....	8
三、冲突定位：三条驱动在什么地方不能同时为真.....	9
四、共同假定，与推翻它的那件材料.....	10
五、溯改、溯改率与那张辨别格.....	12
六、把三条驱动接成一个循环：轮次结构.....	14
七、一次已经发生过的检验：把首次作答变成可撤回，会怎么样.....	15
八、在十一个行当里兑现.....	16
九、判据、读法与不得越过的界线.....	17
十、与最接近的几种说法的分界.....	19
十一、与同一栏目里几篇的分界.....	21
十二、三家为什么各自到不了这一条.....	22
十三、一处强制的不一致：后果效度与可比性不能同时以现在的形式成立.....	23
十四、三家反过来给本文加的三条约束.....	24
十五、不适用的边界.....	25
十六、证伪条件.....	25
十七、反噬预言.....	26
十八、自反：本文自己在哪一格.....	26
十九、结论，与一条写死日期的赌注.....	27
注释.....	28

论课堂中溯改通道的规范性关闭——发起权为何不换手，以及可问责与学习为何要求相反方向的驱动

摘要

课堂上真正的学习为何很少发生，现有三种互不相容的回答：教学路径不对、认知条件不允许、局部改良被制度吸收。本文主张三者都不是主因。课堂之所以长期停在低价值状态，是因为它被规范地做成了一个**只许单向传导的系统**：规则改变教学、教学改变表现，而反方向的每一次改变——由这一届学生实际做出来的东西去修改判据本身——都被现行规范判为不公平、污染或漂移，因而被记为负项。本文把这一类反方向的改变命名为**溯改**，给出五条判定条件与一个可事后清点的比率读数**溯改率**（最近若干轮中，因本轮实际产物而修改了上一层规则的轮次占比），并论证：判据不因产物而改动的体系，其起始端永远停在同一处；溯改为零时发起权不换手，系统在第一次选定的状态上被锁住，而这种锁死不产生任何失效信号，指标反而持续改善。推翻上述共有前提的材料来自被引来作类比的化学自己——同一分子的一个更稳定晶型一旦在某个车间偶然成核，晶种即改写了全球范围的生产条件，原来那条路径此后再也走不通；化学把它记为污染事故，而它在形式上正是产物反过来改写了驱动它的条件。判据只问文件、不问人的判断：**这一轮结束之后，评分规则、任务清单或考纲里，有没有哪一条是因为这一轮学生实际做出来的东西而被改掉的；改了几条，改动记在谁名下。** 本文以中国某试点城市二〇一五至二〇二〇年的一次真实改革为检验：该改革把首次作答变成可撤回（两次考试取高分），而四年追踪显示考试频次、教师中心讲授与机械记忆全部上升，项目式活动被放弃——可撤回性若不伴随溯改，只增加轮次，不改变状态。结论是一条本文不打算缓和的话：可问责要求规则先于且独立于产物，学习要求产物改写规则；课堂被要求同时是两者，而它一次只能是其中一个。本文与同日同栏另两篇处理的是同一问题的不同环节——一篇管一轮里带不过来的状态，一篇管验收落在哪一拍上，本文管轮与轮之间带不回去的规则；三者可以独立为零，正文第十二节对两篇各给一条可现在就做的双向对照。

关键词：溯改；溯改率；发起权换手；动力学捕获；层级单向性；可问责悖论；阶段律

Counted as a Deficit

The Normative Closure of Upstream Amendment in the Classroom: Why the Locus of Initiation Never Changes Hands, and Why Accountability and Learning Require Opposite Directions of Drive

Abstract. Why is genuine learning rare in classrooms? Three incompatible answers dominate: wrong instructional path, prohibitive cognitive conditions, and local improvements absorbed by the institutional field. This paper argues that none is the principal cause. Classrooms remain in a low-value state because they are normatively constructed as strictly unidirectional

systems: rules shape teaching, teaching shapes performance, while every change in the reverse direction—an amendment of the criterion itself occasioned by what this cohort actually produced—is classified by current norms as unfairness, contamination, or drift, and therefore entered as a deficit. This reverse change is here named upstream amendment. Five identifying conditions and a countable ratio (upstream amendment ratio: the proportion of recent rounds in which the governing rule was amended because of that round's actual output) are specified. Write-back is the sole source of any change in the locus of initiation; where it is zero, the locus never changes hands and the system is kinetically trapped in its first-selected state—a trapping that emits no failure signal while indicators continue to improve. The premise shared by all three answers is overturned by material belonging to chemistry itself: once a more stable polymorph of ritonavir nucleated in a single facility, seeding rewrote production conditions worldwide and the original path became permanently unavailable. Chemistry records this as a contamination incident; formally it is a product rewriting the conditions that drove it. The paper's criterion contains no modal terms: after this round, was any line in the grading rules, task list, or syllabus altered because of what these students actually produced; how many lines, and in whose name? A four-year longitudinal study of a Chinese reform that made first attempts revocable serves as a decisive test: revocability without upstream amendment increases throughput without changing state. Accountability requires the rule to precede and be independent of the product; learning requires the product to amend the rule. A classroom is asked to be both, and can only be one at a time.

Keywords: upstream amendment; upstream amendment ratio; locus of initiation; kinetic trapping; hierarchical unidirectionality; accountability paradox; rule of stages

一、导论：三个不该同时为真的判断

关于"课堂上真正的学习很少发生", 眼下有三种回答, 各有充分的经验支持, 而它们不能同时为真。

第一种说: 问题出在教学路径。学生学不到东西, 是因为帮助给错了地方、给错了时候, 或者因为评价把教学逼成了应试。这一支最有力的证据是评价的反拨——改考试, 教学就变; 不改考试, 怎么改教法都会被压回去。它的处方是改评价: 改题型、改频次、改口径。

第二种说: 问题出在结果与路径的错配。一个体系里同时存在若干条通往不同结果的路径, 实际走通的那一条不由哪条最好决定, 而由哪条最容易起步决定。最容易起步的常常不是最稳的那一个。这一支的处方是改起步条件, 不是改目标。

第三种说：问题出在场域。任何在一门课、一个班、一所学校里做出来的局部改良，最终都会被更大的评鉴、认证与排名体系吸收，变成一份可申报的材料。这一支的处方是别在局部使劲，要么改场域，要么承认改不动。

三种回答在处方上互相取消。第一种要改评价，第二种说改评价改的是障碍高度与时间长度，而这两个量在某些条件下根本不进入选择；第三种说不管改什么，改完都会被吸收，因此前两种都是在被吸收的那一层里使劲。反过来，第一种会说第三种把改革失败归给一个谁也够不着的东西，等于取消问题；第二种会说前两种都假定了目标是给定的、只是路径难走，而实际上决定结局的是最初那一步走向了哪儿。

更要紧的是，三家各自每天在做的事，正是另一家诊断为病根的那件事。改评价是第一家的日课，而第三家把“用评价去驱动教学”直接把它列为脱耦的来路；讲究起步条件是第二家的日课，而第一家会说那不过是把难度调低的另一种说法；承认场域不可改是第三家的克制，而第二家会说那是把一个动力学问题当成了热力学问题——认定系统会走向最稳定的那个状态，只是慢一点。

本文不打算在三者之间判谁对。判谁对需要它们站在同一根轴上，而它们不在。本文要做的是找出一条它们**共同假定、而谁也没有说出口**的东西，并用其中一家自己已经掌握、已经发表、只是没有往这个方向用的材料把它推翻。

那条共同假定是：**在一个课堂里，谁驱动谁，是这类系统的性质决定的、在整段过程里保持不变的；因此改善的办法是找准那条驱动，从它下手。**

三家争的是那条驱动是哪一条。三家都没有争的是：驱动的方向是不是固定的。

推翻它的材料在第四节给出，来自化学。而由此得到的判断，三家都不会同意，但三家的证据合起来只能得出它：

> 课堂上真正的学习之所以少，不是因为哪条驱动选错了，是因为课堂被做成一个**只许一个方向传导**的系统。规则改教学、教学改表现，这个方向畅通无阻；而反方向——这一届学生实际做出来的东西去修改判据本身——的每一次发生，都会在现行规范里被记成一个负项。**没有反方向，就没有下一轮从哪一端开始的改变；发起权不换手，系统就停在它第一次选定的那个状态上。它不是没有在变，它是在同一个状态里越长越大。**

这条判断的可查形式很短，全文都可以拿它去问：**这一轮结束之后，评分规则、任务清单或考纲里，有没有哪一条是因为这一轮学生实际做出来的东西而被改掉的？改了几条？改动记在谁名下？**

本文的路线是：第二节交代三家各自的最新立场与它们各自留下的一处松动；第三节把三家的冲突写成一个可判定的形状；第四节找出共同假定并给出推翻它的那件材料；第五、六节建立本文的读数与轮次结构；第七节用一次已经发生过的真实改革做检验，那次改革的结果同时削弱了本文最自然的一条处方；第八节在十一个行当里兑现；第九节给判据与不得越过的伦理界线；第十至十二节与最接近的既有说法、与本栏已发的几篇、以及与三家各自的理由划清界线；第十三节写下三家反过来给本文加的三条约束；余下各节交出边界、证伪条件、反噬预言、自反定位与一条写死日期的赌注。

二、三家的最新立场，以及各自留下的那一处松动

2.1 教育学：评价单独决定学生怎么学

这一支最硬的经验基础是反拨研究。自从一九九三年那篇《反拨存在吗？》把这个领域从思辨推向实证以来，三十余年的积累给出的图景是：高利害考试确实改变教学，而且改变的强度在教学内容上远大于在教学方法上。[1] 这一支的强版本主张是单因的：**决定学生怎么学的是评价制度；教学法只是因变量**。因此任何不改评价而只改教法的努力，最多在一轮之内有效。

时序读数这一支给得很清楚：评价规则先动；教师的任务设计在数周至一个学期内跟上；学生的学习方式在一到两个学期内跟上。跟不上的那一步也很清楚——教学方法。

这一支自己留下的松动，恰恰就在这一步。同一批文献自己确立了两件事：其一，反拨对内容的影响远强于对方法的影响；其二，教师的专业背景与信念会调制反拨的强弱与方向，以至于在同一场考试改革下，不同学校的课堂活动差异并不由考试设计解释，而由教师本身的专业性与信念解释。[2] 换句话说，这一支自己承认：它主张的那条驱动，在传到方法层时会被别的东西挡住或改写。它没有往下问一句：那个挡住它的东西，是不是有时反过来在驱动它。

2.2 化学：走通哪条路，由起步条件单独决定

一八九七年提出的阶段律说：一个体系若有若干个亚稳状态可走，它会依次经过与当前状态最接近的那一个，而不是一步到位走到最稳定的那个。这条经验规律被引用了一百多年，覆盖无机物、金属、金属有机框架、有机小分子、生物大分子、药物、高分子与二维材料。[3]

而它现在正处在一场活的争论里。二〇二三年的一篇再检验用经典成核与生长理论重算了多晶型化合物的间歇结晶，结论是阶段律作为一条普遍规律站不住：经典理论给出的不是顺序过程，而是**所有亚稳相并行成核与生长**；实际观察到的不是亚稳相依次出现，而是**亚稳相按其亚稳程度依次消失**。只有当各相的成核与生长动力学参数恰好被拉开时，体系才在表观上显得像是顺序的。[4] 与之对立的一方则试图为阶段律给出结构与热力学上的说明。[5]

无论哪一方赢，这一支的单因主张都成立而且更强了：**结晶出来的是哪一相，由成核与生长的动力学以及体系当下的条件单独决定；热力学稳定性不是决定项**。时序读数是：过饱和度等条件先动；相在秒到分钟量级出现；向最稳相的转化可以推迟数小时、数十年，甚至因动力学捕获永不发生。跟不上的那一步是最后一步——转化本身。

这一支自己留下的松动在第四节展开，因为它同时是本文推翻共有前提的那件材料。这里先记一句：这一支自己知道，产物一旦大量存在，它会通过晶种反过来改写下一批的条件，使原来那条路径不再可达。它把这件事记在“污染”与“工艺事故”名下。

2.3 管理学：场域标准由日常实践改写

制度理论内部同一年出现的两篇文章，把因果箭头指向相反的方向。一篇说：社会理性化的扩张——管理科学的影响、可问责与透明的要求——从外部改变了组织的环境，于是组织被迫在越来越多的领域里对齐政策与实践；由此产生的不是老式的"说一套做一套"，而是一种更普遍也更严重的形态：**手段与目的之间的脱耦**——政策被认真执行了，执行的手段与它宣称的目的却没有关系。[6] 另一篇说：场域层的逻辑变迁并不起源于场域，而**起源于个体在日常工作中为把事情做完而做出的即兴处置**；这些处置先在组织内被证成、被扩散，然后改写场域。[7]

本文取后一支作为管理学这一家的立场，因为它是三家里唯一一个把因果起点放在下游产物上的：**决定一个做法能否存活的是场域条件，而场域条件是被大量组织日复一日的实践改写出来的**。时序读数是：实践先动；组织内部的正当化跟上；场域标准最后跟上，量级是年。跟不上的一步是最后一步——一旦场域标准跟上了，它就回过头来规定什么算合格的实践。

这一支自己留下的松动是它明写的：它把自己定位为对"制度创业"叙事的一种反例，即变迁并不总是由意图的行动者发起。但它同时承认，这条链条要走通，必须经过组织内部的正当化环节；而正当化用的语言与判据，正是场域给的。也就是说，它主张的那条驱动，其成立条件由它所要驱动的那样东西提供。

2.4 三家的动力式并排

把三家各自写成"什么与什么不容，逼动了什么改变"，形状立刻清楚：

	什么与什么不容	逼动了什么改变	先动的是	跟不上的一步
教育学	可见的成绩表现 × 制度条件（升学、问责）	教与学的路径改道	评价规则	教学方法
化学	成核与生长路径 × 体系当下条件	出现的是哪一相	条件（过饱和度、杂质、界面）	向最稳相的转化
管理学	已做出来的实践 × 既有的形式流程	场域条件本身改变	日常实践	场域标准的追认

三家各锁一条，被逼动的那一样互不相同：一家逼动路径，一家逼动结果，一家逼动条件场。这不是三个视角看同一件事，是三条互斥的单因主张——每一家都认为只有自己那条是决定性的，另两样是被决定的结果。

三家都没有做的，是第三步。

三、冲突定位：三条驱动在什么地方不能同时为真

三家的分歧不是"哪个因素更重要"。若只是权重之争，做一次方差分解就化解了。分歧在于：三家各自主张自己那条是**决定性的**，另两样是被决定的结果。这使它们在同一个具体处境里给出互相取消的处置。

取一个具体处境：一门课，学生当场都能做对，一周后重测大面积失守，换个表述的同类题几乎全错。

- 教育学这一家说：评价考的是当场做对，所以教出来的就是当场做对。改评价——考延迟的、考迁移的——教学自然改道。
- 化学这一家说：改评价改的是障碍高度与时间长度，而在首次作答不可逆的条件下，这两个量根本不进入路径选择。学生第一次交出来的那个做法一旦被记分，它就成了这门课的种子；此后所有的努力都在这颗种子长成的那一相里进行。要改，得改起步条件，不是改目标。
- 管理学这一家说：前两家说的都会发生一次，然后停下。因为无论改评价还是改起步条件，改出来的东西要在这所学校活下去，就必须被写成一份可申报、可评鉴、可与外校比较的材料；而一旦被写成那份材料，它就变成了新的既定流程，下一届又从零开始。

三条不能同时为真。若教育学对，则改评价之后教学方法应当持续改变，而不是变一次就重新固着；若化学对，则改评价这类只动难度与时间的干预不应产生任何持久变化；若管理学对，则改起步条件同样会被吸收，因为它一样要被写成材料。

而三家的证据合起来，恰好指向一件三家都没有说的事。 教育学的纵向研究显示，反拨对方法的影响弱于对内容的影响，且四年之后应试策略的强度回落、而"以考试要求指导教学计划"这一条稳住了。化学的再检验显示，实际观察到的顺序不是好东西依次出现，而是**别的可能性依次消失**。管理学的实践驱动研究显示，日常处置能改写场域，但必须先经过一道用场域的语言写成的正当化。

三条证据说的是同一件事的三面：**在这三个体系里，改变都发生了，而改变的发起处从来没有换过人。** 教育里永远是评价先动；化学里永远是条件先动；管理里那条自下而上的链条，其正当化的判据仍然由场域给。

三家的争论因此可以写成一句话：**三家争的是"发起权"该归谁。** 而三家共同假定的，是发起权是一样有归属的东西——它归谁，是这类系统的性质决定的，在过程里不变。

四、共同假定，与推翻它的那件材料

4.1 把共同假定说出来

把上一节末尾那句话摊开，共同假定是这一条：

> **在一个体系里，谁驱动谁是给定的。这一轮里被驱动而改变的那一样，就是这一轮的产物；反过来去改动驱动方的那一类事件，不是这个体系的正常运行。**

把这条念给三家听，三家都会说"这还用说吗"。教育学不会说评价是被学生表现驱动的——那叫标准不统一。化学不会把"这一批产物改写了下一批的条件"写进机理式的右边——那叫批次污染。管理学承认实践能改写场域，但它把这个过程叫"变迁"，即一种偶发的、需要专门解释的事件，而不是每一轮的常规环节。

三家都没有争论这一条，因为三家都站在它上面。

4.2 推翻它的那件材料：来自化学自己

一九九六年上市的一种抗逆转录病毒药物，其胶囊配方基于当时已知的唯一晶型。上市两年后，为生产该晶型而设计的工艺开始产出另一种更稳定的晶型，溶出度与生物利用度随之失效。真正要紧的不是这件事本身，而是随后发生的事：一旦更稳定的那一晶型的晶种在该公司任何一处实验室或生产车间形成，原来的晶型就再也做不出来了。产品必须撤市重新配方。[8]

在化学自己的账本上，这是一次污染事故，处置办法是清场、换厂房、隔离设备。它被写在"工艺开发的后期风险"名下，回答的是"如何控制多晶型以保证生产可重复"这个问题。

而把它剥到形式层看——去掉全部化学名词，只留下什么对什么做了什么，导致什么——它说的是：

> 一次产出改写了产出它的那套条件，使原来那条路径此后在全局范围内不再可达。

这正是共有前提所排除的那一类事件。而且它不是罕例：同一支文献把"消失的晶型"作为一个持续存在的类现象反复综述，并指出这类案例多次成为专利诉讼的核心争点——法律要判定的正是"原来那条路径究竟还走不走得通"。[9]

于是共有前提的三个部件被同时打掉：

1. 谁驱动谁不是给定的。在同一个体系、同一段时间里，条件驱动产物，产物也可以驱动条件；发生哪一种，取决于这一轮谁先动。2. 被驱动而改变的那一样不是这一轮的全部产物。那一次产出的最重要后果不是产出本身，而是它对条件的改写；而这项后果在任何一张产率表上都不出现。3. 反过来改动驱动方不是异常。它是这一类体系的常规环节，只是被记在了别的名下——负的那一栏。

第三点是本文的入口。

4.3 三家的账本上，这一类事件都在负栏

把同一个形式结构拿到三家的账本上逐一比对，结果整齐得可疑：

反方向的那类事件叫什么	记在哪一栏
化学 产物改写了下一批的条件	批次污染 / 工艺事故（负）
教育学 评分口径因这一届学生的实际表现被改	标准不一致 / 不可比 / 放水（负）
管理学 验收标准因这一次的结果被改	标准漂移 / 移动球门（负）

这不是"记录得少"。三家都有字段，都能查到，都有专门的名称与专门的处置程序。问题在于**符号是负的**。一类事件被系统地记为负值，等于它作为一样正的东西不存在——没有人会去优化它，没有人会去申报它，没有任何一份年度报告会把它列为成绩。

本文把这一类事件命名为溯改，并主张它是这三个体系里唯一能改变"下一轮从哪一端开始"的环节。

五、溯改、溯改率与那张辨别格

5.1 一次溯改的五条判定条件

不是所有的规则改动都算溯改。要算，五条必须同时成立，缺一条不算：

1. **改的是判据，不是执行。** 被改的必须是"什么算做对"这一层的成文物：评分细则、验收标准、考纲、工艺放行条件。改教法、改讲义、改工序参数不算——那是被驱动方内部的调整。
2. **触发者是本轮的实际产物。** 改动的依据必须是这一轮真实做出来的东西，不是上级文件、不是同行做法、不是新理论。
3. **改动可追溯。** 存在一份能指到具体轮次、具体产物的改动记录。口头承认不算。
4. **改动约束下一轮。** 改完之后，下一轮的判定按新规则走。只在本轮酌情放宽而下一轮恢复原状的，叫例外处理，不叫溯改。
5. **改动被记名。** 存在一个可查的责任人或责任位置。匿名的、无人负责的改动在制度上等同于漂移，无法与失控区分，因此不计入。

第五条是最容易被跳过的一条，也是最重要的一条。它是把溯改与"标准不稳"分开的唯一装置。**没有记名的溯改就是漂移；有记名的漂移就是溯改。** 两者在文本上完全一样，差别只在那一栏依据与那一个名字。

5.2 溯改率

＞ 溯改率 $W = \text{最近 } N \text{ 轮中，发生了至少一次溯改的轮次数，除以 } N$ 。

这个读数在三个领域里量纲不同而比率相同：

- **教育：**最近 N 次课程结束后，考纲、评分细则或任务清单被改的次数占比；改动的依据栏须能追溯到具体一届学生实际交出来的东西。
- **化学：**最近 N 批生产中，因本批产物性状而修改了下一批工艺放行条件的批次占比；改动记录在批记录与工艺变更单里。
- **管理：**最近 N 次项目复盘，因本次结果而修改了验收标准或考核办法的次数占比；改动记录在制度文本的版本差异里。

它满足四条性质：

其一，无量纲，三个领域可以并排比较。**其二，可事后清点——**所需的原始记录在多数机构的既有档案里已经存在，不必新建采集系统：教务系统里有大纲修订记录，质量体系里有变更单，项目管理里有版本库。**其三，它把一个印象变成一个数——**"这个组织学不学习"这句话此前只能靠感觉，现在是一个分数。

其四，也是最容易漏又最要命的一条：它与既有的主要指标正交。 举一个既有指标最好看、而这个数最难看的真实形状：一门通过率长期在九成五以上、评教接近满分、教师全部拿过教学奖、而考纲十年一字未改的课程。按现行的任何一套教学质量指标，它是模范；按 W 看，它是零。反过来，一门 W 很高的课程完全可能教得一塌糊涂——那时高 W 读出来的是标准不稳。 **W 不是一个越高越好的分数，它是一个方向读数：它读的是这个系统上一次改变是从哪一端发起的。**

5.3 辨别格：溯改方向 × 轮次通量

单有一个比率还只是一个名字。把它升成两根轴，四格才分得开那些从外面看一模一样的课堂：

	轮次通量低（这一段时间里发生的作答、 批改、反馈次数少）	轮次通量高（次数多）
有溯改（ $W>0$ ，判 据因本轮产物被改 且记名）	甲格·慢改 少而深的作品课、师徒制、临 床带教。改变慢，但每一次改变都改到判 据层。风险是样本太少，改动可能只反映 个别学生。	乙格·循环 事故复盘驱动的训练体系、 成熟的代码评审、迭代式产品团队。发 起处频繁换手，是唯一能持续跃迁的一 格。
无溯改（ $W=0$ ）	丙格·冻结 十年不变的考纲加上低强度教 学。它至少诚实：谁都看得出这里什么也 没发生。	丁格·捕获（本文要打的一格） 高频作 答、密集批改、精细反馈、指标持续改 善——而判据一次未动。

丁格是本文要打的那一格，它也是唯一需要一个框架才看得见的一格。其余三格不需要本文：慢改看得出慢，冻结看得出冻，循环看得出好。

丁格具备三条签名，缺一条则说明这根轴选错了：

1. 它在短期指标上表现为改善。作答量、正确率、参与度、满意度、通过率全部上升，而且是真的上升，不是造假。2. 它的失效不产生任何信号。因为整套记录只记产物，不记状态；判据没有改动，就没有任何一份文件会提到判据本该改而没改。3. 它自我加固。越正规化越严重：评价越严密、越可问责、越追求可比，溯改通道关得越死。一所管理松散的学校反而更可能出现未记名的溯改；一所管理精良的学校连这个漏洞也补上了。

第三条决定了丁格不是失职的产物，而是尽职的产物。这一点是本文与所有“教师不负责任”“管理不到位”式解释的分界，也是本文最不肯让步的一处。

5.4 为什么“改而不变”：动力学捕获

把化学那一支的结论拿到这里，可以说清楚一件反直觉的事：一个体系可以每天都在剧烈变化，而它所处的状态从不改变。

结晶时先出现的不是最稳定的那一相，而是最容易起步的那一相。至于最稳的那一相什么时候出现——按争论中较强的那一方的算法，实际上根本不是“依次出现”，而是别的相依次消失；只有当各相的动力学参数被拉开时，表观上才像顺序过程。[4] 更要紧的是：向最稳相的转化并不保证发生。若转化的通道本身被堵住，体系可以在亚稳相里停留任意长的时间，同时在这一相内部不断长大、变得更完美、缺陷更少、粒径更均匀。

一门课在丁格里的样子完全一样。学生每周作答、教师每周批改、平均分逐年上升、试卷分析报告一年比一年细致。这些全都是在同一相里长大。判据没有改过，所以这门课要长出来的那个东西是什么，从第一年起就定死了；此后所有的努力都在把那个东西做得更完美。

而通道为什么被堵住，第七节会给出一个具体案例。这里先给出机制上的答案：**转化需要新相成核，而新相成核需要一次判据改动；判据改动被记为负项，所以成核事件被系统性地压制。**这不是比方。把"新相"读作"另一套关于什么算做对的规定"，把"成核"读作"这套规定第一次被写进一份约束下一轮的成文物"，整个机制逐字对得上。

六、把三条驱动接成一个循环：轮次结构

三家各给了一条驱动，而三家都只写到第二步就停了：某两样不容，逼动第三样改变。第三步没有人写：**改变之后，它回过头改了什么。**

补上第三步，三条驱动就不是三支互不相干的箭头，而是一个循环。

- **教育学那一条**：可见表现 × 制度条件 → 教学路径改道 → 改道后的教学生产出新的可见表现 → 这些新表现若被写进评分口径，下一轮由表现先动；若没有被写进，下一轮仍由评价先动。
- **化学那一条**：成核路径 × 当下条件 → 出现的是哪一相 → 这一相通过品种、模板与局部浓度改写了条件 → 下一轮由产物先动；若产物被彻底清除、器皿被清场，下一轮仍由条件先动。
- **管理学那一条**：已做出来的实践 × 既有流程 → 场域标准改变 → 新的场域标准重新规定什么算合格的实践 → 下一轮由场域先动；若新标准未被写成成文物，下一轮仍由实践先动。

三条的第三步形状完全一样，而且都取决于同一件事：**那次改变有没有被写进一份约束下一轮的成文物。**写进去了，发起权换手；没写进去，发起权留在原处。

于是本文的核心机制可以压成一句：

> **判据一次也不因产物而改动的体系，其起始端永远停在同一处；改变起始端的，只有溯改这一件事。**

这句话把一个静态的分类问题变成一个可查的时序问题。它给出的预测不是相关，而是**轮次**：

预测一（零溯改课堂的定常形态）：在 $W \approx 0$ 的课程里，每一轮的发起处都是评价；教学路径在二至六周内跟上；学生表现在一个学期内跟上；然后回到起点。四年之后，应试策略的强度会回落（因为教师已经熟练，不再需要那么用力），而"以考试要求指导教学计划"这一条会稳住甚至上升。**指标改善与状态不变同时出现。**

预测二（有效干预的第一个信号）：一次真正起作用的干预，其第一个可观察信号不是成绩变化，也不是满意度变化，而是**发起处换人**——即某一次判据改动的依据栏里，第一次出现"本届学生实际做出来的东西"。这个信号出现的时间比迁移测验上的差异**早一到两个轮次**。

预测二是本文最难被巧合满足的一条。相关关系很容易被巧合满足；顺序关系难得多；而"某一份成文物的依据栏里出现了一类以前从未出现过的依据"是一个离散的、可以指到具体日期与具体文件的事件。

预测三（无效干预的签名）：若一项干预使成绩显著改善而发起处始终未换人，则该改善会在干预停止后一至两轮内消失，且消失时不产生任何信号——因为整套记录里没有任何一栏会记下"判据本该改而没改"。

这三条预测在第十五节被写成可证伪的形式。

七、一次已经发生过的检验：把首次作答变成可撤回，会怎么样

本文最自然的一条处方是：既然体系被锁在第一次选定的那一相里，那就让第一次选择变得可以撤回。这条处方听上去无懈可击，成本也近乎为零。

而它已经在一个足够大的规模上被试过了，结果与预期相反。

二〇一六年起，中国某试点城市把一门高利害升学考试由一年一次改为一年两次，考生可取两次中的高分用于录取。这项改革的**明写目标就是降低考试焦虑与"教到考试为止"的危害**。[10] 就本文关心的那根轴而言，它做的正是把首次作答变成可撤回：第一次考砸了不算数，第二次可以覆盖。

一项覆盖改革前（二〇一五）、改革当年（二〇一六）与四年后（二〇二〇）三个时点、共计五百八十二名高三英语教师的追踪研究给出了结果。[10]

改革当年：

- 高三英语考试频次显著上升——每两周考一次的教师由改革前的一成一升到接近三成七；
- 以考试要求作为教学大纲首要依据的教师，由三成八升到接近五成八；
- 布置机械记忆作业的比例由半成升到近一成四；
- 以教师为中心的讲授方式由二成四升到接近三成九；
- 教师放弃了海报制作、辩论一类项目式活动，换成应试策略；
- 教师的工作压力显著上升，对该项改革的态度显著转向怀疑。

四年之后：

- 应试策略的强度与工作压力都有所回落；
- 而"以考试要求指导教学计划"这一条稳住甚至更高；
- 教师对该政策的态度由怀疑转为**谨慎接受**。

这组数据的形状与本文第六节的预测一逐条对上：轮次通量大幅上升（考试频次翻了三倍），指标层面的适应完成（压力回落、态度转正），而系统所处的状态没有变——甚至比改革前更靠向应试那一相。

关键在于这次改革把可撤回性给在了哪一层。 它给在了**分数层**：哪一次成绩算数，学生可以选。它没有给在**判据层**：什么算做对，一字未改。学生获得的是"重来一次的机会"，而不是"我这一次做出来的东西可以改变什么算做对"。

于是发生的事情正是本文机制所预言的：**可撤回性若不伴随溯改，只增加轮次，不改变状态。**多给一次机会，等于把同一相里的生长再跑一遍；跑两遍的结果是这一相长得更大、更完美、更稳固。这也解释了为什么教师的压力会先升后降而态度会由怀疑转为接受——他们不是被说服了，是**适应了同一相里的更高通量。**

这一节同时做了两件本文不愿回避的事。

第一，它**削弱了本文最自然的处方。**本文原本会写下一句更强的话：让第一次选择可以撤回，锁死就会解除。这句话现在写不下去了。可撤回性是必要的松动，但它作用在被驱动的那一端；只有当被撤回的那一次实际产物**回过头改了判据**，路径选择才真正改变。本文因此把处方收窄为一句：**可撤回性必须与溯改成对出现，单独给出的可撤回性是通量装置，不是路径装置。**

第二，它**给出了一条本文自己的检验。**若上述解释成立，那么在同一场改革里，那些同时改动了校本评分口径（例如把两次考试之间的过程性作品写进评价依据）的学校，应当出现与整体趋势相反的走向；而那些只执行了两次考试、评分口径一字未改的学校，应当完整复现上述全部数据。这份对比所需的资料——校本评价方案的历年版本——在多数学校的档案里已经存在。

八、在十一个行当里兑现

一条判断若只在提出它的那个场景里成立，它就是一个说法。下面十一处不是类比，是兑现：每一处都能据此预测一件具体的、可以去查的事。

一·**大学有备案考纲的必修课。**W 通常为零，且零得有制度理由：改考纲要走教务备案，改了就与往届不可比，与兄弟院校也不可比。预测：这类课程的教学质量指标与学生迁移表现之间的相关性，会随课程"成熟度"上升而下降——越成熟的课越是丁格。

二·**住院医师床边教学。**W 非零但不入账。主治医生会因为这一台手术改下一次的教学重点，这是每天都在发生的溯改；而病历系统与教学档案里没有这个字段，改动不记在任何人名下。按第五条判定条件，它在制度上等同于漂移。预测：一旦这类改动被要求记录，最先发生的不是溯改增加，而是溯改减少——因为现在它要走一道手续。

三·**飞行模拟机复训。**这是 W 最高的一个行当，而且是制度性地高：事件报告与飞行数据监控构成强制反馈回路，训练大纲的修订以具体事件为依据、有编号、有责任部门。预测：航空训练体系中判据改动的依据栏里"具体事件"的出现率，显著高于任何一个教育行当；且这一差异不能由资源投入解释——它由记账规则解释。

四·**开源项目的代码评审。**W 极高且全部留痕：一次讨论可以直接修改贡献者指南，且改动带提交记录与作者。这是乙格的教科书形态。预测：在同一个组织里，代码库的判据改动率会显著高于该组织人事考核办法的改动率，尽管两者面对的不确定性相当。

五·手工艺师徒。 W 非零但不可分辨——因为判据从来没有成文，溯改与随意在文本上无法区分。这是本文读数的一处真实边界：**W 只有在有文判据的体系里可读。** 预测：这类体系的能力传承质量与任何一种外部评估之间几乎不相关，而这并非评估做得不好，是被评估的东西没有可比对象。

六·企业新人引导。 W 非零而分层：新人的反馈经常改文档，很少改验收标准。溯改发生在被驱动的那一层，没有到判据层。预测：新人引导材料的年改动率远高于试用期考核办法的年改动率，且两者的比值可作为该组织学习结构的读数。

七·临床指南更新。 这是一个中间形态：指南确实因新证据而改，但证据的产出被上一版指南规定的终点指标塑形。溯改存在，而它绕了一大圈，轮次以年计。预测：那些终点指标本身被改过的领域，其指南的实质性修订率高于那些只改阈值不改终点的领域。

八·品种选育。 田间实际表现改写选育目标，这是农业里最常规的溯改，而且记名（品种审定材料上有）。预测：把选育目标冻结、只优化育种效率的项目，其品种在真实生产条件下的表现会先于试验田表现出退化。

九·学术同行评审。 W 近乎为零，且是本文最尴尬的一个案例：审稿意见改论文，几乎从不改投稿须知。判据由编辑部与学科共同体持有，稿件是产物；产物改写判据的通道在制度上存在（专刊、社论、法学声明），但被使用的频率极低。预测：一本刊物的投稿须知年改动率，与该刊所发论文的方法学多样性正相关，而与影响因子不相关。

十·事故复盘。 军事与工业的复盘制度是溯改的成熟形态：这一次的实际结果改写下一次的作业规程，有编号有责任人。预测：复盘制度的有效性差异，主要由“复盘结论是否进入作业规程”决定，而不由复盘会开得是否深入决定；只出报告不改规程的组织，其同类事故复发率与不做复盘的组织没有显著差异。

十一·人工智能产品的验收。 这是眼下最尖锐的一处。模型的实际产出改写评测集，是这个行当每天在做的事；而它同时被指为数据污染。**同一个动作，在一个方向上叫迭代，在另一个方向上叫作弊，分界线取决于改动有没有被记名与公开。** 这一处最清楚地展示了本文的机制：溯改与污染在文本上是同一件事，全部差别在那一栏依据与那一个名字。

十一处里，第三、四、八、十四处的 W 制度性地高，而它们有一个共同点：**这些行当的失败会以外人可读的形式出现**——飞机掉下来、软件崩溃、庄稼减产、事故重演。W 高不是因为这些行当更开明，是因为它们的判据被外部现实持续证伪，因而不得不改。**教育行当的判据几乎不会被外部现实以可读的形式证伪**，这是它的 W 结构性偏低的根本原因，也是任何一份“向航空业学习”的建议在教育里落不了地的原因。

九、判据、读法与不得越过的界线

9.1 那一句问话

本文的判据不问人的判断，只问文件——它可以直接拿去问教务系统、变更单与版本库：

> 这一轮结束之后，评分规则、任务清单或考纲里，有没有哪一条是因为这一轮学生实际做出来的东西而被改掉的？改了几条？改动记在谁名下？

把它拿到六个具体场景里跑一遍，六个答案互不相同，其中两个是查出来的而不是想出来的：

场景	答案
大学有备案考纲的必修课	零。且零有制度理由——改了就与往届不可比
住院医师床边教学	非零，但无记录、无记名，在制度上不可与漂移区分
飞行模拟机复训	制度性非零，改动有编号、有依据事件、有责任部门（查出来的）
手工艺师徒	不可读——判据从未成文，溯改与随意无法区分
企业新人引导	非零，但改的是文档不是验收标准；溯改止于被驱动的那一层
开源代码评审	非零且高，改动带提交记录与作者（查出来的）

飞行训练与开源评审这两处，不能从本文的命题直接推出来——它们是去查了才知道的，而且正是它们让"教育行当为什么不同"这个问题有了答案（见第八节末尾）。

9.2 三种分层引用

本文的主张分三层，各自可以单独被引用、单独被推翻：

- **第一层（最强、最易倒）**：驱动结构判断——溯改被记为负项，因而发起权不换手，因而系统被捕获在第一次选定的状态上。
- **第二层**：溯改率 W 作为跨领域可比的读数，以及那张两轴四格的辨别格。
- **第三层（最稳、最便宜）**：一条不依赖前两层的经验主张——在同一制度环境内的一组同类课程中，其规则改动的依据栏里出现"本届学生实际产物"的课程，学生的迁移表现更好。这一条只需要教务系统里的大纲修订记录与一次迁移测验，不需要接受本文的任何概念。

第一层被推翻时，后两层仍然站得住。本文明确写下这一点，既是诚实，也是保护。

9.3 伦理界线：三条写死的规定

W 是一个比率，比率几乎必然会被拿去考核人。因此三条必须写死，缺一条本文不可被用于任何考核：

1. **W 指的是位置，不是人。** 它读的是"这个岗位、这门课有没有一条溯改通道"，不是"这位教师称不称职"。一门课 W 为零，绝大多数情况下说明的是这所学校的备案流程，不是这位教师的态度。把 W 挂到个人绩效上，是对本文的直接误用。
2. **不得为了把这个数做好看，而在安全关键的体系里制造判据变动。** 麻醉、飞行、核电、药品放行的规则稳定性本身是安全资产。这些领域的 W 合理值低，且每一次溯改必须走正式的变更管控制程序。在这些地方，慢就是对的。
3. **已经按 W 做出不利安排的机构，必须公开编码规则并提供复核通道。** 什么算一次溯改、依据栏怎么算追溯得上、谁来判，这些必须是公开可查的成文规则；被判定的一方必须能提出复核。

三条之外还有一句，它属于下一节：一个读数一旦进入考核，最省事的达标方式通常是在文书上做手脚。这一条的具体形态见第十七节，本文找不到出口。

十、与最接近的几种说法的分界

本文的判断很容易被读成某个既有概念的改名。下面逐一交代，每一条给出判定形式：若观察到甲，则本文对而该概念错；若观察到乙，则反之。

其一·双环学习（Argyris & Schön 1978）。这是最近的邻居，近到必须先说清楚。它说到哪一步：单环学习在既定目标下纠错，双环学习修改目标本身；组织之所以停在单环，是因为行动者的防卫性推理与组织的防卫性惯例。分离线在**病灶的位置**：双环学习把它放在行动者的心理与组织文化，处方是提高心理安全感、推行探询式对话；本文把它放在**账本的符号**——溯改被记为负项，这是一条会计规则，与行动者的心理无关。判定形式：若在一个心理安全感量表得分很高、探询文化良好的团队里，W 仍为零，且零的原因可追溯到"改了就不可比、不可问责"这条明文规则，则本文对而双环学习错；若在同样的成文规则约束下，提高心理安全感能把 W 抬起来，则反之。

其二·古德哈特定律与坎贝尔定律。它说到哪一步：一个度量一旦成为目标，就不再是好度量；社会指标越被用于决策，越会扭曲它所监测的过程。分离线：古德哈特讲**度量被操纵**——数字上去了、实质下来了；本文讲**度量不被修改**——数字和实质可以同时上去，而系统仍锁在低价值状态。判定形式：若能找到一个不存在任何操纵、指标与实质同步改善、而迁移表现长期不变的体系，则本文对而古德哈特无话可说；若所有这类体系一经细查都存在指标操纵，则本文被古德哈特吸收。顺带说明：两者可以同时为假——一门课可以既无操纵又零溯改。

其三·成功陷阱、能力刚性与内卷化。把本文的命题剥到只剩形式，是"一个方向被禁，体系停在初始状态"。这与"越成功越失败"那一族反转模板形状不同，但相近到必须点名。成功陷阱说的是成功的正反馈挤出探索，是**资源配置**问题；能力刚性说的是核心能力同时是核心僵化；内卷化说的是精细化而无质变。判定形式：若在一个长期失败、从无成功可陷、资源也从未被成功占用的体系里（例如一门通过率长期偏低、无人愿教的课），W 仍为零且原因同为不可比性，则本文对而成功陷阱错。这一条是本文与该族最硬的分界：**成功陷阱预测不出"失败体系同样不动"这一格**。内卷化最贴的一处是"同一相里越长越大"这个形象，但它描述状态，本文给出的是发起处这个可查的位置读数。

其四·反拨效应（Alderson & Wall 1993 以降）。它说到哪一步：高利害考试确实改变教学，改内容强于改方法，且受教师信念、学校类型、社会压力调制。分离线：反拨研究的整个设计把考试当作自变量、教学当作因变量，因此它问的是"这个方向有多强"；本文问的是"另一个方向为什么不存在"。判定形式：**反拨研究预测"改考试则教学变"**；本文预测**"改考试而不开溯改，教学变一次之后重新固着，第二轮起发起处仍是考试"**。这两条在现有的纵向反拨数据上就可以分出胜负——第七节已经用了一份。

其五·形成性评价（Black & Wiliam 1998; Bennett 2011）。它说到哪一步：依据学生表现调整教学，可以提高学习。这听上去就是溯改，因此必须划得最细。分离线：形成性评价的溯改止于**教学路径**，不及于**判据本身**；它的全部文献都预设标准不动，否则"学生达到了标准"这句话就没有内容。而这一支自

己的批评性检讨已经指出，这个词至今不代表一组界定清晰的做法或器物，各种实现之间差异极大，效果因而应当预期为高度不齐。[11] 判定形式：**若一所学校大规模实施形成性评价而考纲与评分细则的年改动率仍为零，则本文对而"形成性评价即溯改"错。**所需资料就在教务档案里，是本文最便宜的一条检验。

其六·制度变迁的形态分类 (Streeck & Thelen 2005)。这是本文必须交代的方法学邻居——因为本文的操作（清点规则文本的改动、追溯其依据）正是这一支跑熟了几十年的做法。它说到哪一步：制度变迁有叠层、转换、漂移、替代四种形态，通过追踪规则文本的改动与其实际用途来分类。分离线：这一支分类的是**改动的形态**（谁加了一层、谁改了用途），本文读的是**改动的依据来源**（这次改动是被上游发起的还是被下游产物发起的）。最像的一格是漂移。判定形式：**漂移预测"规则不变而效果变坏"；本文的丁格预测"规则不变而效果稳定甚至改善"。**一个指标持续向好、判据一字未动的体系，在漂移的分类里没有位置，而它正是本文要打的那一格。

其七·必要变异度定律 (Ashby 1956)。外圈的一位，也是最容易被误当成本文的一位：控制器的变异度必须不小于被控体系的变异度，否则控不住。分离线：这条定律讲的是控制器**不够复杂**，处方是增加控制器的变异度；本文讲的是反向通道**被规范禁止**，与复杂度无关。判定形式：**若把评分细则做得极其精细（控制器变异度大增）而 W 仍为零、迁移表现仍不变，则本文对而必要变异度在此无解释力。**眼下几乎所有的评价标准精细化改革都在提供这个观察。

其八·手段一目的脱耦 (Bromley & Powell 2012)。它说到哪一步：在可问责与透明压力扩张的环境里，老式的"说一套做一套"变得难以为继，取而代之的是政策被认真执行、而手段与目的失去关系。[6] 分离线与判定形式：脱耦讲的是**手段与目的之间断开**，本文讲的是**产物与判据之间的反向通道被关**。二者可以独立发生：一个手段与目的高度耦合、人人都在做真正该做的事的体系，同样可以 $W=0$ 。若能找到这样一个体系（例如一门目标明确、手段有效、师生皆认真、而考纲十年不改的课），本文对而脱耦无话可说。反过来，若所有 $W=0$ 的体系一经细查都存在手段一目的断开，本文被脱耦吸收。

最近的一位是双环学习。它仍不能吸收本文，理由只有一条：双环学习把"行动者有能力修改目标"当作给定，因此看不见"修改目标在制度上被记为负项"。一旦承认这一点，它的处方就塌了——心理安全感改不了会计规则。

上述八种说法的共同盲区

把上述八种说法逐个问一句"它把什么当作给定，因此看不见什么"，答案有一个交集：

- 双环学习把"行动者有能力修改目标"当作给定，看不见修改目标被记为负项；
- 古德哈特把"度量与被度量者是两样东西"当作给定，看不见度量规则本身是这个体系的一个组分；
- 反拨研究把"考试在教学之外、先于教学"当作给定，看不见考试也是这一轮的产物之一；
- 形成性评价把"标准是给定的、学生向标准移动"当作给定，看不见标准可以向学生移动；
- 制度变迁分类把"变迁的发起者在制度之外"当作给定，看不见变迁可以由制度自己的下游产物发起。

交集是同一块地：规则与它所规制的东西属于两个层级，而层级之间只有一个方向是正当的。

所有人都站在这块地上，包括互相批判的各派——因为可问责性本身就建立在它上面。如果规则可以被它所规制的产物改写，那么“依规则问责”这句话就失去了意义。这不是某一派的疏忽，是全部人共同的立足处；也正因为共同，没有人能回头看它。

十一、与同一栏目里几篇的分界

本栏已有六篇触到相邻的轴，其中三篇用的正是同样这三个学科，两篇与本篇同日发表、同问一个问题。同日问题同三学科是读者一眼就能认出的自撞，不指名是不诚实的；而三篇既然共处一栏，彼此就是对方最近的一位，划得清才谈得上各自成立。

其一 · 《方程两边都有它》（同为教育学、化学、管理学）。那一篇问的是一个东西：一次课堂轮次里，那个被记入学习者账户而未进入净变化的成分该怎么认出来；它的答案是一个辨别维度（旁观轮次），它要打的那一格是“分岔的那一步被代做而即时产物合格”。本篇问的是一个驱动：为什么整个体系长期停在低价值状态；答案是一个轮次结构，本篇要打的是“三条链路都在正常运转、指标全好、而判据一字未动”那一格。判定形式：在一门完全没有代做、没有提示、学生全程独立走完每一个分岔步的课程里，那一篇预测不存在旁观轮次因而学习应当发生；若这样的课程 $W=0$ 而迁移表现仍差，则本篇对而那一篇为空。反之，一门考纲每年因学情而改（ W 高）、而分岔步仍被系统性代做的课程，那一篇对而本篇为空。两篇互补而不重叠：那一篇管一次轮次内部的归属，本篇管轮次之间的发起权；叠加时流失速度不是相加而是相乘——分岔步被代做而溯改为零，既学不到又改不了。

同时诚实记下一处继承：本篇第五节末尾关于“路径选择取决于可逆性”的直觉，与那一篇的第三处升级同源，那一笔记在那一篇账上。本篇把它从单次作答的可撤回推到判据的可修改——同一条化学原理在两个层级上的两次使用；而第七节的那次真实改革恰好证明，这两个层级不能互相顶替。

其二 · 《每一轮都从新鲜表面开始》（同日同栏，同为教育学、化学、管理学，同问这个问题）。这是本篇最近的一位，近到必须放在第二条。那一篇问的与本篇是同一个问题，用的是同样这三门学科，找到的共同假定也在同一处——发起处在轮与轮之间不动。分歧在它被什么固定住。那一篇把病灶放在轮次内部的结算规则：每一轮真实生成、真实占位的未完成中间体，在结算时被强制清零，因而下一轮只能从起始态开始；它的读数是结转率，问的是“下一轮的头三分之一在处理上一轮留下的哪一处未完成”。本篇把病灶放在轮与轮之间的判据修改权：产物改判据这一类事件被记为负项，因而下一轮的发起处仍是判据那一端；读数是溯改率，问的是“判据里有哪一条因这一轮的产物被改掉”。

一个管带过来的状态，一个管带回去的规则；两者可以独立为零。判定形式：一门每节课都从上一节留下的具体未完成处接着往下走（结转率高）、而考纲十年一字未改（溯改率零）的课程，若其迁移表现仍不改善，则本篇对而那一篇为空；反之，一门考纲每年因学情而改（溯改率高）、而每节课仍从复习旧知开始、当堂检测全对、进度严格按教案推进（结转率零）的课程，若其迁移表现仍不改善，则那一篇对而本篇为空。两项在同一份教务档案里都可以查，因此这条分界不是修辞，是一次可以现在就做的对照。

两篇的关系是互补而非重叠，且叠加时不是相加：结转为零使每一轮回到同一个起点，溯改为零使每一轮走向同一个终点。起点与终点同时被钉死，中间那段路无论怎么修都通向原处——这是两篇合起来才给

得出的一句话，任何一篇单独说不出。加上下一条那一篇，三者合起来是同一个循环的三处断口：起点带不过来、验收采错了拍、终点改不动；**任修其一，另两处都会把它压回去，这正是课堂改革逐项试过、逐项被吸收的形式解释。** 诚实记下一处：本篇第六节把三条驱动接成循环的那个动作，与那一篇是同一个动作；本篇不主张在先后上谁更承重。

其三 · 《先动的总是同一边》（同日同栏，同为教育学、化学、管理学，同问这一个问题）。 那一篇把病灶放在**节拍**：一堂课里三条驱动的周期不同——教师侧以分钟计、学习者侧以天到周计、制度侧以学期到年计——而课堂验收对准了最快的那一条，于是每次验收都落在“教师已改讲法、学生表现已跟上、而学生自己还没有开始改自己的做法”那个时刻；它的读数是节拍比与发起处轮转率。本篇把病灶放在**判据的修改权**：无论验收落在哪一拍，产物改判据这一类事件都被记为负项；读数是溯改率。

一个管验收落在哪一拍上，一个管验收之后判据动不动。二者可以独立发生，判定形式因而是双向的：若一门课程把验收节拍调到学习者侧那一条（节拍比接近一，那一篇预测被压掉的那条驱动重新入账），而其考纲与评分细则仍一字未改、迁移表现仍不改善，则本篇对而那一篇为空；反之，一门考纲每年因学情而改（溯改率高）、而全部验收仍落在教师改讲法那一拍上、迁移表现仍不改善，则那一篇对而本篇为空。两处调整互不依赖：改节拍不必动判据，动判据不必改节拍。

还有一处必须让出去：那一篇用的推翻材料是异相成核——有现成界面时体系不再走离自己最近的那一步，改走那个结构邀请它走的一步。**这一条比本篇所用的品种污染更靠前**：它说的是发起处会易主，而本篇说的是易主要靠一次记名的判据改动才落得下来。**本篇不主张自己那一条更承重；本篇只主张，没有那次落下来的改动，易主不会跨到下一轮。**

其四 · 《缝合致死》。 那一篇的病灶是“把活的东西外化并固定”这个动作本身，预测的是成功导致死亡。本篇的病灶是反向通道被记为负项，预测的是**成功与失败都不导致改变**。判定形式：一个从未成功过、从未被度量过、也从未把任何东西固定成可脱离情境的对象的实践（那一篇预测它不会因缝合而死），若其 $W=0$ 而同样停滞，则本篇对而那一篇为空。

其五 · 《收得回的那一遍》。 那一篇的可撤回落在单次作答的执行层，调节的是可读性。本篇的溯改落在判据层，调节的是判据本身。第七节已经用一次真实改革把两者分开：分数层的可撤回不产生判据层的溯改。

其六 · 《那一段谁走的》。 那一篇按“有没有产出”定义它的对象。本篇的溯改恰恰要求有产出，并且产物改写了规则——在定义上互斥，两篇不可能落在同一格。

十二、三家为什么各自到不了这一条

三家停在原地，不是因为疏忽。是因为各自的核心问题恰好使这一步成为不必要的。

教育学到不了，因为它的问题是“如何设计一场好考试”。这个问题的形状要求考试先于教学、独立于教学——否则效度就无从定义。一场随本届学生表现而改的考试，在它的框架里不是好考试，是失效的考试。它看不见溯改，不是因为没看，是因为它一旦承认溯改是常规环节，效度理论的地基就要重铺。

化学到不了，因为它的问题是"如何可重复地得到我要的那一相"。可重复性要求产物不改写条件——产物一旦改写条件，工艺就不可重复了。所以它必须把这类事件识别出来并**消灭**它：清场、换厂房、隔离设备。它是三家里唯一把这类事件研究得最透的，也正因为研究得透，它必须把它记在负栏。**它对这件事的全部知识都是为了防止它发生。**

管理学到不了，因为它的问题是"场域逻辑如何变迁"。变迁在它的框架里是一个需要解释的事件，因此它要找的是那些**发生了变迁**的案例。而本文关心的恰恰是那些**什么都没发生**的场合——发起权始终不换手的那种稳态。这类稳态在变迁研究里不是数据，是背景。它**看不见溯改为零的意义**，因为零在它的抽样框里不出现。

三条合起来说明一件事：**这一条不是三家的综合，是三家各自的问题结构所排除的那一格。**

十三、一处强制的不一致：后果效度与可比性不能同时以现在的形式成立

上一节说三家各自到不了这一条。这一节要说的更重一点：三家里有一家，在不承认本文这个环节的情况下，它自己内部就不自洽。

教育测量学同时持有两条，两条都是它的地基，而且都写在教科书的前几章。

甲条 · 可比性。 一个分数要能被用来做决定，就必须与别人的分数、与往届的分数可比。这要求判据先于表现、独立于表现：命题的双向细目表先定，评分细则先定，等值与标准设定先做。若判据随本届学生的表现而改，等值就无从做起，跨届比较就失去意义，任何一次不利判定都无法在申诉中辩护。**这一条是整套考试制度的合法性来源。**

乙条 · 后果进入效度。 自二十世纪八十年代末以来，效度的主流定义把评价使用的**后果**纳入效度论证：一场考试的效度证据不只包括它测到了什么，还包括它被这样使用之后产生了什么。这一条是现代效度理论最重要的一次扩张，也是"评价的社会后果不是外部议题"这句话的全部根据。

两条在通常情形下相安无事。**它们在一种情形下正面相撞：当本届学生的实际产物构成了"这套判据正在系统性地生产错东西"的证据时。**

按乙条，这是效度证据，而且是最直接的一类效度证据——它正是"这样使用之后产生了什么"。既然是效度证据，且指向不利，效度论证就要求判据被修改。按甲条，因本届产物而修改判据，恰恰是不可比性，是这套制度最不能允许的动作。

同一份材料，按乙条要求改判据，按甲条禁止改判据。

眼下的处置办法是层级隔离：把这类材料从效度论证里挪出去，交给"考试的社会影响""政策后果"这类外部议题，由政策而非由测量来处理。这个处置在文本上有效，代价是把乙条掏空——**一个永远不会导致判据改动的后果，不是效度证据，是背景说明。**

于是这一支面临一个二选一：

- **要么放弃乙条**，把后果退回到效度之外，承认效度只管测到了什么。那是回到二十世纪八十年代之前，且它自己已经明确拒绝过。
- **要么承认溯改是效度论证的一个环节**——即承认存在一类正当的、记名的、入账的判据改动，其依据是本届产物。而一旦承认这一点，**甲条就不能再以现在的形式成立**：可比性不再是判据不动，而是"判据改动的规则本身可比"。

这不是本文替它选。这是本文指出它必须选，而它现在没有选。

同一个动作可以对人工智能教育产品的验收再走一遍。这个行当同时持有：其一，模型的实际产出应当用来改进评测（否则评测会过时）；其二，用模型产出去改评测集是数据污染（否则分数不可信）。两条同样在一种情形下相撞——当产出显示评测集本身在测错东西时。它现在的处置同样是层级隔离：把前者叫"迭代"，把后者叫"泄漏"，而两者的分界线在实践中只能靠**改动有没有被记名与公开**来划。这恰好就是本文第五节的第五条判定条件。

本文已经给自己留了退路：这一处的强制不一致，可以被"咬牙接受乙条被掏空"化解——即公开承认后果只是背景说明。那样它不是不一致，只是弱了很多。这条退路被写进第十六节的证伪条件第九条。

十四、三家反过来给本文加的三条约束

碰上有用的对手，代价是它们也会削掉你原本想说的话。三处，逐条写清削掉了什么。

其一 · 化学削掉了本文的价值排序。 本文原本会写下一句更强的话：**溯改率越高越好，应当尽可能打开溯改通道。** 化学把这句话取消了。工业上大量的结晶刻意跑在亚稳相上——巧克力要的是特定的那一种脂肪晶型，许多药物特意选用溶出更快的亚稳晶型，最稳定的那一相不是想要的那个。被捕获在哪一相是好是坏，取决于要的是哪一相。于是本文的主张从"应当提高 W"收窄为：**应当能读出 W，并知道自己当前锁在哪一相；锁死本身不是罪名，不知道自己锁在哪里才是。** 这一处动的是本文的价值排序，不是补一个条件。

其二 · 管理学削掉了本文的适用范围。 规则的稳定性本身是可问责性的来源。一个 W 很高的体系在法律上可能是不可问责的：标准随时在变，任何一次不利判定都无法辩护。于是本文的适用范围从"所有教学场合"收窄为"**已经决定要产生迁移性学习的课程**"。岗前合规培训、执业资格考试、驾驶执照考核的目的恰恰是可比性与可辩护性，那里的 W 应当低，而且低是对的。**这一处动的是适用范围本身。**

其三 · 教育学削掉了本文最省事的实现方式。 那一支自己的检讨已经指出，一个概念一旦被要求在制度上执行与记录，它会退化为一批可申报的器物而不是一组做法。[11] 这直接毁掉了本文最自然的处方——"要求每门课记录溯改"。这一条的完整形态在第十七节。

十五、不适用的边界

四处明写：

其一·判据未成文的体系读不出 W。 师徒制、家庭教养、非正式共同体里，溯改与随意在文本上无法区分。本文的读数在这些地方无效，不要硬套。

其二·安全关键体系的合理 W 值低。 见第九节第二条。在这些地方拿本文去催促判据变动是危险的误用。

其三·本文不处理能力不足型的停滞。 一个教师不会教、学生基础差、资源匮乏的课堂，同样会停滞，而那是另一种病。本文要打的那一格明确要求短期指标在改善——**不改善的那些，先去查别的。**

其四·高 W 不等于好。 高 W 可能读出的是标准不稳、管理失控、或者迎合。W 是一个方向读数，不是一个质量分数；单独看它会得出荒谬的结论。要判断一个体系是不是在跃迁，必须同时看第五节那两根轴。

十六、证伪条件

先写出最强的竞争解释，那句最朴素、最难反驳的话：“**其实不就是应试教育吗？其实不就是评价压力太大、大家没时间搞别的吗？**”若真是如此，那么决定迁移表现的应当是评价压力的绝对强度，而不是驱动结构；把压力降下来，学习就该发生。第七节那次改革恰恰是一次降压尝试，而结果相反——但那只是一个案例，不足以排除该解释。

因此本文的核心证伪条件写成这个形状：

＞ **控制评价压力强度、任务认知负荷与机构层合规压力之后，若课程层的溯改率 W 与学生迁移表现之间不存在剂量—反应关系，则本文的驱动结构判断为伪；**届时应归因于评价压力的绝对强度而非驱动方向。

样本纪律写死：**同一所大学同一学院的八至十二门同类必修课**，在学时、班额、生源、教师职称构成上同质，自变量（历年大纲与评分细则的改动记录及其依据栏）有真实变异。**不得拿两个国家或两个地区做对比充数**——那种比较两边处处不同，所谓“控制掉某变量”在方法上站不住，只能作启发式说明。

另有八条，其中第三条最便宜，所需数据在多数教务系统里已经存在：

1. **顺序检验：**若在一项有效干预中，迁移表现的差异出现**早于或同时于**发起处换人（即判据改动的依据栏第一次出现本届学生产物），则第六节的预测二为伪。
2. **无效干预的签名：**若一项使成绩显著改善而发起处始终未换人的干预，其效果在停止后两轮内仍然保持，则预测三为伪。
3. **最便宜的一条（第三层主张）：**在同一学院的一组同类课程中，若“大纲改动的依据栏出现本届学生产物”与学生迁移表现之间无关联，则第九节第三层主张为伪。所需资料：教务系统的大纲修订记录 + 一次延迟迁移测验。
4. **形成性评价的分离：**若一所大规模实施形成性评价的学校，其考纲与评分细则的年改动率显著高于对照校，则本文与形成性评价的分离线不成立。
5. **心理机制的分离：**若在成文规则不变的前提下，仅提高团队心理安

全感即可显著抬高 W ，则本文对双环学习的分离线不成立。6. **失败体系的一格**：若长期低通过率、无人愿教的课程其 W 显著高于模范课程，则本文与成功陷阱一族的分界不成立。7. **可撤回性的分层**：在第七节那场改革中，若那些同时改动了校本评分口径的学校与只执行两次考试的学校，其教学行为数据走向一致，则第七节的解释为伪。8. **记名条件的必要性**：若未记名的判据改动与记名的判据改动在下一轮的约束力上没有差别，则第五节第五条判定条件多余， W 的定义须重写。9. **强制不一致的退路**：若教育测量学公开承认后果只是背景说明、并把它退回效度论证之外，则第十三节所指的不一致不成立——那时它只是弱了，不是自相矛盾，本文对它的那一处主张作废。

若本文被证伪，我们会学到什么：若 W 与迁移表现确实无关，那么"改变的发起处"这个位置就不是一个有用的读数，教育治理应当回到强度与配置这一层——而这将是一个正面的结论，因为它会把大量资源从"打开通道"这类结构改造上收回来，投向更直接的地方。

十七、反噬预言

本文一旦被制度采用，最省事的实现方式是：**要求每门课每学期提交一份《依据学情调整记录表》**。

于是溯改变成一项文书， W 变成一个申报数字，而实际的判据改动仍然不发生——甚至更少，因为现在改动要多走一道申报手续。第八节第二处已经预言了这个形态：一旦床边教学的即兴调整被要求记录，最先发生的不是溯改增加而是溯改减少。

比这更省事的一种是：把《依据学情调整记录表》做成一个模板，让每门课都填得满满的，然后年度报告里写"全院溯改率百分之百"。填表的人不必改任何判据，因为表格问的是"你有没有依据学情调整"，而这个问题的答案永远是"有"。

唯一能想到的减缓办法是：**只读已生效的判据文本的版本差异，不读申报表；且要求每一处差异都指到具体的产物与具体的日期**。但这个办法自己也有出口——判据文本可以被写得足够模糊，模糊到无须修改即可涵盖一切；于是版本差异永远为零，而没有人违规。

我看不到出口。任何使 W 可核查的装置，都同时使 W 可申报。这一条写在这里，不是待办事项，是本文交出去的一个洞。

配套的一条底线：**验收必须看已生效判据文本的实际差异，不看是否设有溯改制度**。有制度而零差异，与没有制度，在本文的读数上是同一格。

十八、自反：本文自己在哪一格

本文提出的判据，对本文自己适用吗？适用，而且结果不好看。

本文是一篇论文。一篇论文的文本在发表之后是冻结的。读者读了本文之后所做的任何改变，都不会溯改到本文——因为学术出版制度把"因读者反应而修改已发表文本"这件事记为负项：那叫勘误，严重的叫撤稿。**本文的溯改率结构性地为零。**

后果具体有三条：

第一，本文不能靠被讨论来改进，只能靠被使用来检验。 讨论会产生引用，引用不会改动本文的任何一行。因此本文把最便宜的那一条检验（第十六节第三条）单独标出来，就是为了让"使用"这条路比"讨论"那条路更近。

第二，本文这套读数一旦被采用，第一个被读出 $W=0$ 的对象就是学术论文本身，包括本文。 第八节第九处已经把这一格写进去了：审稿意见改论文，几乎从不改投稿须知。本文没有理由认为自己例外。

第三，本文写死了一条有日期的赌注（见下节），正是为了给自己制造一次强制溯改。 若到期未命中，本文的第一层主张就有一个公开的、不可撤销的负项。这是一篇冻结的文本能给自己安排的唯一一次判据改动。

还有一处更难看的：**本文自己就在丁格里。** 本文写得越细致、兑现处越多、划界越清楚，它在"论文质量"这套指标上就越好看；而这些全部是在同一相里长大。判据——什么算一篇有贡献的论文——本文一条也没有改动，也没有能力改动。

十九、结论，与一条写死日期的赌注

课堂上真正的学习之所以很少发生，不是教学路径选错了，不是认知条件不允许，也不是局部改良被场域吸收。是因为课堂被规范地做成了一个只许单向传导的系统：**判据改变教学，教学改变表现，而表现改变判据的每一次发生，都被记在不公平、污染或漂移的名下。** 记为负项的那一类改变，正是唯一能让下一轮换一个地方先动的那一类改变。没有它，系统就在第一次选定的那个状态上越长越大，而所有的指标都在改善。

这条判断的根，比教育更深一层：**可问责要求规则先于并独立于它所规制的产物；学习要求产物改写规则。** 二者要求相反方向的驱动。一间教室被同时要求做到两件事，而它一次只能做到其中一件。这不是一个可以靠更好的设计消解的张力，只能靠**在时间上把两者分开**——明确哪些轮次归可问责，哪些轮次归溯改，并把后者也记名、也入账、也算成绩。

赌注，写死日期：

> **到二〇三一年十二月三十一日为止，至少有三所大学（或同等规模的教学机构）在其课程大纲修订的正式流程里，设立"本次修订的依据须注明本届学生实际产物"这一必填项，并公开发布年度溯改率。**

写死什么不算命中：

1. 仅要求教师提交"学情分析"或"教学反思"一类文书； 2. 仅在质量报告里自述"依据学情调整"而无可核对的判据文本差异； 3. 仅在选修课或试点班实施，未进入正式的大纲备案流程； 4. 与其他改革打包实施，无法分离； 5. 只公布调整次数，不公布调整的依据来源； 6. 设立了通道而年度实际使用为零——**有制度而零差异，与没有制度是同一格。**

最后交出两个本文解释不了的洞。其一：第五节第五条要求溯改必须记名，而第八节第二处的床边教学表明，大量真实的溯改恰恰发生在无人记名的地方；本文的读数在这些地方读出零，而那个零是错的，本文没有办法修正它。其二：第七节的那次改革给出了一个太干净的解释，而四年之间同时还有别的事情在变；本文无法排除那些变化，只能等那份校际对比资料。

注释

[1] 反拨研究以 Alderson & Wall (1993) 为转折点；关于"强度在内容上大于在方法上"的读数，见 Alderson & Hamp-Lyons (1996) 与后续综述。

[2] 关于课堂活动差异不由考试设计解释而由教师专业性与信念解释，见 Green (2006) 的课堂观察研究；关于教师因素对反拨的调制，见 Watanabe (2004)。

[3] 阶段律的原始提法见 Ostwald (1897)；其跨材料类别的引用范围，见注 [4] 所引文献的综述部分。

[4] Cardew (2023) 用经典成核与生长理论重算多晶型化合物的间歇结晶，指出经典理论给出的是各亚稳相**并行**成核与生长，实际观察到的是各相按亚稳程度**依次消失**。

[5] Threlfall (2003) 尝试为阶段律给出结构与热力学上的说明。两方的对立至今未决。

[6] Bromley & Powell (2012)。该文把"手段一目的脱耦"与传统的"政策—实践脱耦"分开，并把它归因于可问责与透明压力的扩张。

[7] Smets, Morris & Greenwood (2012)。该文与注 [6] 同年发表，因果方向相反，构成管理学内部一处未消解的对立。

[8] 该案例的原始描述见 Bauer 等 (2001) 与 Chemburkar 等 (2000)；"晶种一旦形成，原晶型再也做不出来"这一条，见 Bučar, Lancaster & Bernstein (2015) 与 PNAS (2024) 对该案例的复述。

[9] Bučar, Lancaster & Bernstein (2015) 同时指出，多起此类案例成为专利诉讼的核心争点。

[10] Huang & Zeng (2025)。该研究覆盖二〇一五、二〇一六、二〇二〇三个时点，共五百八十二名高三英语教师；本文第七节所引全部数据均出自该文。改革的明写目标（降低考试焦虑与"教到考试为止"的危害）见该文导论部分对政策文件的转述。

[11] Bennett (2011)。

参考文献

1. Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115–129.
2. Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280–297.
3. Argyris, C., & Schön, D. A. (1978). *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.
4. Ashby, W. R. (1956). *An Introduction to Cybernetics*. London: Chapman & Hall.
5. Bauer, J., Spanton, S., Henry, R., Quick, J., Dziki, W., Porter, W., & Morris, J. (2001). Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research*, 18(6), 859–866.
6. Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
7. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
8. Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *The Academy of Management Annals*, 6(1), 483–530. <https://doi.org/10.1080/19416520.2012.684462>
9. Bučar, D.-K., Lancaster, R. W., & Bernstein, J. (2015). Disappearing polymorphs revisited. *Angewandte Chemie International Edition*, 54(24), 6972–6993. <https://doi.org/10.1002/anie.201410356>
10. Cardew, P. T. (2023). Ostwald rule of stages—Myth or reality? *Crystal Growth & Design*, 23(6), 3958–3969. <https://doi.org/10.1021/acs.cgd.2c00141>
11. Chemburkar, S. R., et al. (2000). Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Organic Process Research & Development*, 4(5), 413–417.
12. Green, A. (2006). Watching for washback: Observing the influence of the IELTS academic writing test in the classroom. *Language Assessment Quarterly*, 3(4), 333–368.
13. Huang, L., & Zeng, J. (2025). Sustainability of washback effects: A longitudinal study of China's twice-yearly NMET reform. *Language Testing in Asia*, 15, 19. <https://doi.org/10.1186/s40468-025-00358-9>
14. Leonard-Barton, D. (1992). Core capabilities and core rigidities: A paradox in managing new product development. *Strategic Management Journal*, 13(S1), 111–125.
15. Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, 14(S2), 95–112.
16. Ostwald, W. (1897). Studien über die Bildung und Umwandlung fester Körper. *Zeitschrift für Physikalische Chemie*, 22, 289–330.
17. Smets, M., Morris, T., & Greenwood, R. (2012). From practice to field: A multilevel model of practice-driven institutional change. *Academy of Management Journal*, 55(4), 877–904. <https://doi.org/10.5465/amj.2010.0013>
18. Streeck, W., & Thelen, K. (Eds.) (2005). *Beyond Continuity: Institutional Change in Advanced Political Economies*. Oxford: Oxford University Press.
19. Threlfall, T. (2003). Structural and thermodynamic explanations of Ostwald's rule. *Organic Process Research & Development*, 7(6), 1017–1027.
20. Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing: Research Contexts and Methods* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum.
21. Wei, S., et al. (2024). Crystal size, shape, and conformational changes drive both the disappearance and reappearance of ritonavir polymorphs in the mill. *PNAS*, 121(15), e2319127121. <https://doi.org/10.1073/pnas.2319127121>

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