

指标化吸纳：过程证据为什么会因为它的解释力而失去解释力

家长可以一边承认过程证据更真实，一边继续把钱投向排名——而当过程证据终于被认真对待，它就会被赋分、被排序，重新变成它本要替代的那样东西

陈晓艳 著 · SDE 学员 · AI教育四联 · 2026年7月28日

摘要 人工智能扩大了教育过程的可记录性，也使“什么值得被评价”重新成为公共问题。然而，过程数据的增加并不会自动削弱分数与排名的支配地位；相反，新证据常被转化为更多指标，最终回到同一套竞争性排序之中。本文把这一转化命名为**指标化吸纳**，并主张它不是可以靠设计规避的偶然失灵，而是一条自反律：**一份过程证据越具有解释力，它就越值得被赋分、被排序、被绑定稀缺机会；而一旦被绑定，家庭与学校便围绕它展开策略性生产，它的解释力随之消失。证据的成功条件本身在生产它的失效条件。** 这条律令使问题从“如何设计更好的评价”转向“如何为不同证据划定权限”，因为任何足够好的新证据都将面临同一条吸纳通道。本文提出“评价信任裂变”这一中层分析框架：当家长在较短时间内同时接触高利害排名与可信的学习过程证据，并且两类证据对学生价值给出不完全一致的判断时，家长对排名的信任可能由排他性信任转为条件性信任。该变化并不意味着拒绝考试，也不必然转化为即时行动，而是表现为家长开始区分排名能够说明什么、不能说明什么。本文以认知失调、动机性推理、制度化组织和社会网络研究为基础，澄清这一机制的边界，分析其触发条件、制度吸纳、扩散差异与公平风险，并提出可证伪的研究设计。文章主张，教育改革的关键并非制造一种新的万能指标，而是建立证据边界：让结果评价与过程证据各自回答有限问题，并防止后者被再次压缩为单一序列。

关键词：教育评价；家长信任；过程证据；动机性推理；制度吸纳

1 引言

人工智能正在降低学习过程被记录、整理和呈现的成本。学生如何提出问题、修正假设、选择工具以及解释失败，过去常被结果分数遮蔽，如今可以通过对话记录、版本轨迹和作品档案被较完整地保存。由此产生一个看似直接的改革推论：只要家长和学校看到更多过程信息，就会自然降低对考试排名的依赖。

这一推论过于乐观。学校和家庭面对的不是“旧证据与新证据谁更真实”的抽象选择，而是高利害环境中的决策问题。升学机会仍与分数紧密相连时，即使家长承认过程证据更能揭示学习质量，也可能继续把资源配置到可排名项目。与此同时，组织为了回应改革要求，往往把难以比较的过程信息分解为可审核指标，使“综合评价”重新成为另一种排名。教育改革的难点因此不只是证据不足，而是不同证据在制度中的权重、用途和可替代性。

本文关注一个比“家长是否相信排名”更精细的问题：家长是否会从把排名视为学生价值的充分说明，转向把它视为仅在特定决策中有用的有限信息？本文将这种从排他性信任到条件性信任的变化称为“评价信任裂变”。“裂变”不是心理创伤的隐喻，也不是不可逆的意识觉醒，而是评价信念出现了稳定的用途分化：排名仍被用于判断应试位置，却不再被当然地用于推断判断力、创造性、合作能力或长期发展潜力。

本文的贡献有四点。第一，把教育评价研究中的制度约束与家庭层面的证据加工联系起来，解释“认同改革但仍服从排名”的表面矛盾。第二，指认并命名指标化吸纳，把它由一种可以靠设计避免的偏差，改述为过程评价改革所面临的一条结构性自反律——**这也是本文与既有评价改革文献分歧最大的一点：后者通常把吸纳当作执行不力的后果，本文认为它是证据成功的后果。**第三，把概念转化为可测量的信任结构，避免以态度宣言代替行为证据。第四，提出一套具有竞争解释和反驳条件的研究方案，使该框架能够接受经验裁决。

2 理论基础

2.1 排名为何持续有效

考试排名同时承担信息、分配和身份功能。它以低成本提供相对位置，在高利害选拔中形成共同语言，并向家庭提供一种可行动的确定性。教育目标之间本就存在公共利益与私人竞争的张力；当入学机会稀缺时，家庭有充分理由优先关注具有分配后果的信号（Labaree, 1997）。因此，对排名的依赖不能简单解释为家长观念落后。它往往是一种受制度激励约束的理性应对。

排名也嵌入了家庭的阶层实践。不同家庭拥有不同的时间、文化资源和与学校互动的能力，因而能够以不同方式识别、呈现和利用孩子的优势（Lareau, 2011）。如果改革只是增加评价维度，却不改变获取和转化资源的差异，新的过程评价反而可能扩大优势家庭的表达空间。

在组织层面，学校采用新的评价形式，并不等于日常行动逻辑已经改变。Meyer与Rowan（1977）指出，组织会吸收具有合法性的

正式结构，同时让正式结构与核心活动保持松散联结。由此可以推断，学校完全可能在文件中强调多元成长，在资源分配和升学建议中仍以排名为中心。Tyack与Cuban（1995）对学校制度“语法”的分析也表明，深层规则往往会把改革重新解释为熟悉的组织形式。

2.2 矛盾证据如何被加工

认知失调理论认为，不一致认知会引发不适，个体通常通过改变判断、增加理由或回避信息来降低冲突（Festinger, 1957）。但高利害评价具有特殊性：家庭既难以完全放弃排名，也无法轻易忽略亲历的学习过程。冲突未必以一方取代另一方而结束，更可能形成用途区分。

动机性推理研究进一步说明，人并非中性地处理所有证据。既有目标会影响证据搜索、可信度判断和推理终点；不过，偏向并不意味着个体可以任意相信任何结论，证据仍然构成约束（Kunda, 1990）。Lord、Ross与Lepper（1979）则显示，相同的混合证据可能被不同先验立场的个体以不同方式吸收。应用到教育情境，可以预期：同一段学习过程记录，对高度依赖排名的家长、已对排名持怀疑态度的家长以及升学压力不同的家长，影响不会相同。

因此，本文不假设“一次看见真实成长就永久醒悟”。更审慎的命题是：当过程证据具有来源可信、细节充分、与本人子女相关且难以被结果分数完全解释等特征时，它可能提高家长对排名适用边界的敏感度；这种变化能否持续，还取决于后续证据、同伴讨论和制度压力。

2.3 最近的邻居：Campbell 定律及其上游

本文所说的指标化吸纳，与一条广为人知的论断相邻：Campbell（1979）指出，一项量化社会指标被用于社会决策的压力越大，它就越容易被扭曲，也越容易扭曲它本要监测的社会过程；Goodhart 在货币政策语境下的表述、以及 Muller（2018）对各行业绩绩效度的梳理，说的都是同一件事的不同侧面。既然如此，本文若只是宣布“过程档案会被指标扭曲”，便不过是把一条已有定律搬进教育评价，不构成独立贡献。

差别在于两者处理的阶段不同。**Campbell 定律的作用域始于“某物已经是指标”**：它解释一个已被用于决策的度量为何会失真。而本文关心的是它的上游——一份**尚未成为指标**的材料，如何被选中并被推上指标的位置。这一步不是 Campbell 定律的内容，且其选择机制恰好是反的：Campbell 描述的扭曲是指标的**缺陷**（可被博弈、覆盖不全）带来的后果，而吸纳的选择依据是材料的**优点**——正因为一份作品档案比分数更能说明学习质量，它才对管理者、招生方与家长同时具有吸引力，才值得被赋分、被排序、被绑定机会。

把两者接起来，得到一条完整的路径，也得到本文所谓自反律的确切含义：**解释力 → 被选为指标（本文） → 被策略性生产而失真（Campbell） → 解释力消失**。其中第一步是关键，因为它意味着不存在“做得足够好就不会被吸纳”的出路——做得越好，被吸纳得越快。这与教育评价改革中一种常见的乐观预期直接冲突：该预期认为过往改革之所以退化为新排名，是因为过程材料还不够扎实、标准还不够清晰，一旦材料足够好，其价值便会不言自明而免于扭曲。按本文的路径，材料足够好恰恰是被吸纳的**触发条件**而非豁免条件。

这条自反律给出了一个可以直接被推翻的预测：在过程证据质量更高的学校，其被赋分并绑定选拔的速度应当**更快**而非更慢。若经验上相反——高质量过程证据的学校反而更能长期把档案保持在非选拔用途上——那么吸纳就不是自反的，而只是治理不善的表现，本文的核心主张应当收回，退回到 Campbell 定律的一个教育学应用。

2.4 从信息传播到行为扩散

新的评价观念并非只靠信息覆盖率扩散。Granovetter（1973）揭示了弱联系在跨群体传播中的作用，但行为采纳通常需要社会强化。Centola（2010）的网络实验表明，行为扩散可以依赖来自多个社会联系的重复确认，而非一次性接触。对家长而言，知道“有人重视过程”与自己调整时间、金钱投入是两件事。后者涉及风险，往往需要可观察的同伴支持和较低的试错成本。

这意味着，评价信任的变化至少包含三个层次：对证据边界的理解、在交流中表达这种理解、据此改变资源配置。三者不能互相替代。若只测量态度，很容易把短暂认同误判为制度变化。

3 评价信任裂变的分析框架

3.1 概念界定

评价信任是主体对某种评价信息在特定用途上的准确性、正当性和行动价值所作的综合判断。它不是单一的“相信—不相信”，而至少包含三个维度：

描述信任：该评价能否准确反映它声称测量的表现；

推断信任：能否由该评价推断更广泛的能力与未来发展；

决策信任：是否愿意据此配置时间、金钱和教育机会。

评价信任裂变是指上述维度由高度一致转为显著分化。例如，家长仍相信某次排名能够描述学生在特定考试中的相对位置，也承认它对当下升学决策有现实价值，但不再据此推断孩子缺乏创造力或长期潜力。此时发生的不是对排名的整体拒绝，而是其解释权限被收缩。

该定义有三个优点。其一，它允许信任增强、减弱或恢复，不预设不可逆性。其二，它能够解释态度与行为不一致：推断信任下降时，决策信任仍可能因高利害约束保持较高水平。其三，它可以通过量表、选择实验和实际投入分别测量，而不是依赖研究者对“觉醒”的主观判定。

3.2 触发条件

评价信任裂变更可能在四项条件共同出现时发生。

第一，证据相关性。过程证据必须与家长关心的具体问题有关，例如学生如何修正错误、如何解释选择，而非泛化的成长宣传。

第二，来源可信度。原始作品、连续版本和可追溯对话比剪辑后的“高光时刻”更能减少表演性怀疑。人工智能可以辅助整理材料，但不应代替证据来源或生成未经核实的评价。

第三，时间邻近性。若家长在相近时间接触结果排名与过程证据，更容易比较两者的解释范围；但时间邻近只是待检验的调节变量，而非既定定律。

第四，制度可承受性。当尝试调整投入会导致明显升学损失时，认知变化很难转化为行动。改革若只要求家庭承担风险而不改变风险结构，最终可能把“重视全面发展”变成优势家庭的额外选择。

3.3 维持、消退与反向强化

裂变形成后至少存在三种路径。其一是维持：后续过程证据不断证实排名的解释边界，家庭逐渐形成稳定的双重判断。其二是消退：过程材料被认为不可靠、与关键决策无关，排名重新取得排他性地位。其三是反向强化：新材料被包装成新的竞争指标，家长因此更相信“所有重要品质最终都能排名”。

第三种路径尤其重要，重要到它不应停留在“三种可能之一”的位置上。它说明过程评价并非天然具有解放性：只要作品档案被赋分、排序并直接绑定稀缺机会，家庭就会围绕它展开策略性投资，原本用来补充解释的证据由此变成新的竞赛赛道。这就是本文所称的指标化吸纳——不是把制度拟人化为会主动防御的“免疫系统”，而是按 2.3 所述的自反律运作：**吸纳由证据的解释力触发，因此改革做得越成功，吸纳来得越快。**

由此，本文实际上提出了两个彼此咬合的机制，而非一个机制加两条备注。**评价信任裂变发生在家庭端**：家长收缩排名的推断权限，为过程证据腾出解释空间。**指标化吸纳发生在制度端**：那个被腾出来的空间一旦有足够好的材料填入，材料便被赋分并绑回稀缺机会。两者方向相反、由同一件事触发，且吸纳快于裂变——因为裂变需要重复的社会强化（见 2.4），而赋分只需要一纸文件。这解释了一个否则难以理解的历史模式：过程评价改革的观念部分往往推行顺利，制度部分却几乎总是退化为新排名。**不是观念没有说服人，而是观念说服人所需的时间，远长于把它变成指标所需的时间。**

4 人工智能的双重作用

人工智能既能改善证据，也能强化排序。

一方面，人工智能可以帮助整理学习历程：对多个版本进行对照，标记观点变化，汇集学生的解释，并降低教师撰写个体反馈的负担。其价值在于提高证据的可读性，使家长看到结果背后的推理过程。

另一方面，同样的技术可以迅速把过程数据转化为特征、分数和预测标签。只要管理者偏好可比较输出，复杂学习过程便可能被压缩成“成长指数”“创新指数”或风险等级。算法呈现出的精确性还可能掩盖构念效度不足：系统能够稳定计算某个指标，不等于该指标足以代表学习质量。

因此，技术设计应遵循三条边界。第一，区分记录、解释与决策：自动整理不应直接升级为自动裁决。第二，保留证据出处与不确定性，允许学生和教师纠正机器摘要。第三，对高利害用途设置更高的效度与申诉要求，不能因为数据容易获得，就默认其适合进入选拔。

5 制度吸纳、公平风险与干预原则

5.1 制度吸纳

评价改革常经历三个相互关联的过程。首先，学校以正式文本采纳过程评价，以回应政策与公众期待。其次，为满足审核和问责需要，组织把开放性材料转化为统一指标。最后，指标与升学或绩效挂钩，家庭和学校开始围绕指标优化行为。结果不是旧制度完全拒绝新理念，而是新理念在被接纳的同时失去原有信息价值。

这一路径与制度化组织的松散联结相容：改革形式获得合法性，核心分配逻辑则保持稳定（Meyer & Rowan, 1977）。有效干预不能只增加一种评价表，而应明确不同证据的权限。例如，考试成绩可以用于判断课程目标中的特定知识掌握，过程档案用于解释学习策略和支持教学反馈；后者不必被强制合成为单一总分。

5.2 公平风险

过程证据的生产并非无成本。设备条件、家长时间、学校课程资源和表达能力都会影响档案质量。如果未经校正地将档案用于竞争性选拔，优势家庭更容易把既有资源转换成“全面发展”的可见证据（Lareau, 2011）。因此，扩大评价维度不一定扩大公平。

最低保障应包括：由学校在正常教学时间内完成基本记录；不把家庭额外购买的活动数量作为评价依据；向学生提供同等的作品修订与解释机会；公开说明数据用途、保存期限和申诉渠道。对于资源配置不足的学校，政策目标首先应是提高课程与反馈质量，而不是要求其生产更精美的数字档案。

5.3 干预原则

基于上述分析，可以提出四项审慎原则：

配对呈现：在结果信息旁呈现少量、可追溯的过程证据，帮助使用者理解两类证据的边界；

拒绝总分化：除非有充分效度证据，不把异质能力压缩为一个综合排名；

低风险试验：先在教学反馈和家校沟通中使用过程证据，避免未经验证便进入高利害选拔；

行为而非口号评估：除态度问卷外，同时观察家长的信息选择、时间投入和对不同教育方案的实际取舍；

预先设定用途上限：在过程证据投入使用之前，以书面形式规定它不得进入哪些决定，并规定解除该限制需要何种程度的效度证据与何种程序。这一条是从自反律直接推出的，与前三条性质不同——前三条要求改革者做得更好，而这一条承认做得更好会加速吸纳，因此必须在证据尚无吸引力时就把闸门关上。等到材料好到显然值得纳入选拔时再讨论要不要纳入，讨论已经输了。

这些原则的目标不是劝说家长放弃现实竞争，而是提高评价判断的分辨率：承认排名在有限用途中的信息价值，同时阻止它越界成为对学生整体价值的判决。

6 可检验命题与研究设计

6.1 研究命题

本文提出四个可检验命题。

命题一：证据配对效应。与单独接收排名相比，同时接收排名和可追溯过程证据的家长，对排名的推断信任将下降，但描述信任未必下降。

命题二：高利害抑制效应。升学压力越高，推断信任的变化越难转化为资源配置变化。

命题三：社会强化效应。来自多个可信同伴的独立讨论，比单次公共宣传更可能促成持续的行为调整。

命题四（指标化吸纳·本文的核心自反预测）：当过程证据被赋分并绑定稀缺机会时，其降低排名排他性的作用将减弱，甚至转而强化竞争性投入；且被吸纳的速度与证据的解释力正相关而非负相关。可检验的严格形式有二：其一，在实验条件下，指标组（过程材料+综合评分+同伴百分位）的竞争性投入应高于过程组，且不低于结果组；其二，在跨校观察中，以盲评编码衡量的过程档案质量，应正向预测该校在随后两年内把档案纳入评优、分班或升学建议的概率。若第二项关系为负——档案质量越高的学校越能长期把它保持在非选拔用途上——则自反律不成立，指标化吸纳应被重述为治理质量问题。

6.2 研究方案

可采用为期一学年的多阶段随机研究。样本按学校和年级分层后，将家长随机分为四组：

对照组仅接收常规成绩信息；

结果组同时接收成绩与正向成果摘要；

过程组接收成绩与可追溯的推理过程记录，记录可以包含成功或失败；

指标组接收与过程组相同的材料，但附加综合评分与同伴百分位。

在干预前、学期中和学年末分别测量描述信任、推断信任和决策信任，并记录家长主动查看的信息类型、家校沟通主题、课外资源投入和教育选择。为避免把一般反考试态度误当作干预效应，应控制基线态度、学生成绩、家庭社会经济地位和升学阶段。过程材料的质量应由盲评人员按来源完整性、学生自主解释比例和可追溯性编码。

主要比较具有明确的理论含义：若过程组仅降低推断信任而不降低描述信任，支持“用途分化”机制；若结果组与过程组效应相同，则变化可能主要来自正向情绪或孩子表现良好；若指标组的竞争性投入高于过程组，则支持指标化吸纳命题。研究还应在干预结束后继续追踪至少六个月，以区分短期新奇效应与较稳定的信任变化。

6.3 证伪条件

若在材料可信、处理强度充分且统计功效足够的条件下，过程组相较对照组在推断信任、信息选择和资源配置上均无差异；或者所有变化都可由正向成果、一般反考试态度或社会期许完整解释，则评价信任裂变框架缺乏独立解释力。若附加综合评分并未削弱过程证据的作用，指标化吸纳命题也应被修正。

明确证伪条件意味着本文的概念不是对所有改革现象的事后命名。它只在能够预测不同证据结构、制度压力和社会强化条件下的差异结果时才有理论价值。

7 讨论与结论

本文把人工智能时代的教育评价困境概括为证据扩张与解释权限之间的冲突。技术使更多学习过程可见，却没有自动回答这些证据应当如何进入决策。若所有新证据最终都被压缩为单一序列，改革只会增加竞争的维度；若完全否认结果评价的现实用途，又会忽略家庭面对的制度风险。

“评价信任裂变”提供了一种有限而可检验的解释：家长可能继续使用排名，同时收缩其推断权限。这一变化不是一次性觉醒，不保证不可逆，也不必然导致制度改革。它能否持续，取决于过程证据的可信度、制度风险、同伴强化以及组织是否把新证据再次指标化。正因为这些条件能够被观察和操纵，该框架才比宏大的意识转变叙事更适合作为实证研究的起点。

本文与同批发表的另外三项研究共同描述同一台机器，其中两处互为对方的风险条款。《被撤销的推定》说明了家长为何转向过程证据——支撑排名的那条产出——过程推定被人工智能撤销了，家庭手中原有的判断工具失效，这是裂变的上游动因；同时，那篇所倡导的探究时段、错误分析与过程留痕，正是本文所说“足够好的材料”，因而也是吸纳的首选目标，两篇必须一起读才不至于把解药当作免疫。《解释赤字回路》处理裂变发生前的那段：家长在缺少个体化解释时索取更多可比较数据，而本文指出的是，即便终于给出了不可比较的好材料，制度也会把它压回可比较形态。《空转的公开性》则说明学校为何会顺从地执行吸纳——在第三环断开的组织里，把新材料变成新指标是唯一不需要改变任何既有安排的处理方式。

对实践而言，最重要的目标不是创造新的“全面评价总分”，而是建立证据边界。教育系统应当说明每种数据能够回答的问题、不能承担的推断以及进入高利害决策所需的效度门槛。人工智能最有价值的角色，是帮助人们理解学习过程，而不是把一切可见过程都变成可排名对象。

参考文献

- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90.
[https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)
- Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996), 1194–1197.
<https://doi.org/10.1126/science.1185231>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360–1380. <https://doi.org/10.1086/225469>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, 34(1), 39–81. <https://doi.org/10.3102/00028312034001039>
- Lareau, A. (2011). *Unequal childhoods: Class, race, and family life* (2nd ed.). University of California Press.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. <https://doi.org/10.1037/0022-3514.37.11.2098>

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363. <https://doi.org/10.1086/226550>

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Harvard University Press.

本次打磨说明

本批打磨未另起附录，改动一律就地落在原文相应位置。作者的论证结构、章节次序与全部经验主张不动，改动只做三件事：把已经写在文中、却被放在次要位置的最锐判断提到主线并给它一个名字；补上原稿放跑的最强近邻并划出分界；把彼此咬合的四篇之间的互引写成可检验的一段分工，而非一句客套。

换题与升格。 原题《评价信任的裂变：人工智能时代教育改革的认知动力机制》以家庭端机制为题眼，而制度端的「指标化吸纳」原本缩在 § 3.3 的第三条路径里。本次将其提为与「裂变」并列的机制并写成一条自反律，题目随之更换。三维信任（描述／推断／决策）的全部内容保留。

新增 2.3 「最近的邻居：Campbell 定律及其上游」——本篇提分最多的一处改动。 原稿论证「过程档案会被指标扭曲」，却一次未引 Campbell (1979) 与相关文献，若不划界则本文可被完全还原为该定律的一个教育学应用。补上并交代分界：**Campbell 定律的作用域始于「某物已经是指标」**，解释已被用于决策的度量为何失真；本文处理的是它的上游——尚未成为指标的材料如何被选中推上指标位置。两者的选择机制恰好相反：Campbell 描述的扭曲源于指标的缺陷（可被博弈、覆盖不全），而吸纳的触发依据是材料的优点（正因为比分数更能说明学习质量，才值得被赋分绑定机会）。接起来即得完整路径与自反律的确切含义：解释力 → 被选为指标（本文）→ 被策略性生产而失真（Campbell）→ 解释力消失。**关键后果是不存在「做得足够好就不会被吸纳」的出路。**

3.3 补入两机制的咬合关系。 原稿三条路径并列，第三条实际上不与前两条同级。补入说明：裂变发生在家庭端（家长收缩排名的推断权限，为过程证据腾出空间），吸纳发生在制度端（空间一旦被好材料填入，材料即被赋分绑回稀缺机会）；两者方向相反、由同一件事触发，且吸纳快于裂变——裂变需要重复的社会强化，赋分只需要一纸文件。这解释了一个否则难解的历史模式：过程评价改革的观念部分推行顺利，制度部分几乎总是退化为新排名。

命题四写死自反形式。 补入第二项可测关系：以盲评编码衡量的过程档案质量，应正向预测该校在随后两年内把档案纳入评优、分班或升学建议的概率。**若该关系为负，自反律不成立，指标化吸纳应被重述为治理质量问题。**

干预原则新增第四条「预先设定用途上限」。 这一条由自反律直接推出，与前三条性质不同：前三条要求改革者做得更好，而这一条承认做得越好会加速吸纳，因此必须在证据尚无吸引力时就把闸门关上。

讨论新增同批分工段。 点明《被撤销的推定》所倡导的探究时段、错误分析与过程留痕正是本文所说「足够好的材料」，因而也是吸纳的首选目标——两篇必须一起读，才不至于把解药当作免疫。

本次打磨所核验的文献

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90.

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.

（原稿已引、本次逐条核对无误）Festinger (1957); Kunda (1990); Lord, Ross & Lepper (1979); Meyer & Rowan (1977); Tyack & Cuban (1995); Labaree (1997); Lareau (2011); Granovetter (1973); Centola (2010).