

# 标记齐全，学习空转

——把教育评价从“标记达标”拉回“过程是否真发生”的操作手册

只讲方法、技术与操作的实践文 · 约 8 千字 · 延伸自《标记空转》 · 德麦国际 Demai International ·  
sdeuniverses.com

一间教室里，可以同时坐着两种学生，而任何一份成绩单都分辨不出他们。一种学生的学习真发生了：他被某道题卡住过，为一个概念反复重来过，某个下午忽然把两件原本不相干的事接上了。另一种学生的学习没有发生，但他的每一格标记都齐全：作业按时交、格式规整、绩点漂亮、查重通过、证书一摞。你翻遍学籍系统，找不到任何一个字段能把这两个人区分开来——因为系统记录的从来不是“学习是否在这个人身上发生过”，而是“有没有一串合规的标记被产出”。

这正是母论文所说的**标记空转**。教育里有一整套认证装置：学分、绩点、证书、评分、查重合格。它由伦理话语（“学术诚信”）、量化考核（分数与排名）、检测技术（查重与监考）三样东西耦合而成，运转起来极其精密，能精确告诉你“谁在何时修了什么、过没过、达没达标、重复率多少”。但它有一个结构性的盲区：它无法判断这些标记背后，是否真发生过一次不可替代的学习过程。它认的是计量终点，不是过程本身。

盲区一旦存在，就会被利用。因为在这套装置面前，**规避标记远比真做便宜、安全、可规模化**：刷绩点比真正搞懂一门课便宜，考证比获得那项能力便宜，把作业做得“合规漂亮”比经历一次真正的思考便宜，让 AI 或基底代笔再改到查重通过，比自己写一遍便宜得多。于是在装置照不到的空转地带，“标记合格但学习没发生”的学生被批量生产出来。学校随后升级检测（更严的查重、AI 检测、更密的监考），学生随后升级规避（改写降重、绕过检测、更精致的应付），这两件事不是对立的，而是同一台装置内部两个组件的互相校准与升级——打假与造假一起把装置越擦越亮，却始终没碰到那个盲区。检测越精密，规避越精致，双方都在进步，唯独“学习有没有发生”这个问题，从头到尾没人回答。

这篇手册不谈教育理念，只给老师和学校一套可操作的东西：怎么识别标记空转、怎么把评价从“标记达标”重新拉回“过程是否真发生”。它的全部方法都围绕母论文的一个判断展开——出路不在于再升级一层检测，而在于坚持**过程主体性**，让“是否真发生过”这件事重新变得可见。下面从看清现象开始。

## 一、先看清：教育里"标记空转"长什么样

要治空转，先得认得出空转。它最迷人的地方在于：现场一切正常，所有标记都在，指标甚至很漂亮。空转不是标记出错，恰恰是标记全对而过程缺席。下面三个场景是它最常见的形态。你不需要额外的仪器，只要换一个观察角度——不看标记本身，而看标记背后那个过程有没有发生。

### 场景一：绩点漂亮，但人是空的

一个学生绩点很高，课门门优秀，奖学金年年拿。可你随便挑一门他拿了高分的课，问他这门课到底解决了什么问题、哪个环节最难、他当时是怎么过去的，他答不上来，或者只能背出考纲上的标准表述。他把每一门课都当成一次"标记采集任务"来完成：摸清评分点、押中考点、把答案对齐给分标准。他非常擅长产出合规标记，甚至比真学的同学更高效——因为他省掉了真学那部分。绩点这个标记完整记录了"他过了、他达标了、他排名靠前"，却一个字都没记录"学习有没有在他身上发生"。这就是母论文说的：**装置精确标记了计量终点，却对终点背后的过程一无所知**。绩点越漂亮，越容易让所有人（包括学生自己）误以为过程一定发生过。

### 场景二：证书满墙，但事情不会做

另一个学生，简历上证书一长串，各类等级考试、技能认证、竞赛奖项齐全。可真把一件对应的事情交给他做——用那门语言完成一次真实沟通、用那项技能解决一个没见过的具体问题——他就卡住了。证书记录的是"他在某个标准化测试里达标了"，而达标可以通过针对考试的专门训练获得：刷题、背模板、掌握应试套路。这套训练和"真正获得这项能力"是两条路，前者短得多、便宜得多、可复制得多。于是墙上挂满了标记，能力却没长出来。证书越多，越像一堵挡在过程前面的墙：它替过程作了证，让人不必再去看过过程本身。这是空转地带最典型的规模化产物——**标记的堆积被误当成能力的堆积**。

### 场景三：查重通过，但没发生过思考

一篇课程论文，查重率压得很低，格式规范，引注齐全，看起来无可挑剔。但通读下来，它没有一处是作者被某个问题真正难住、然后自己想出来的痕迹：观点是拼接的，论证是搬运的，甚至整篇是由 AI 或基底生成、再改写检测到通不过为止。查重这个检测技术，能精确告诉你"这段文字和已有文字的重复率是多少"，但重复率低不等于思考发生过。一个人可以完全不思考，却产出一篇重复率为零的文章——只要他会规避检测。母论文点得很清楚：**检测标记的是"是否与已有物重复"，而不是"是否有过一次不可替代的过程"**。查重通过因此常常成了空转最体面的证书：它把"没有抄"说成了"是我想的"，而这两件事之间隔着整整一个过程。

三个场景形态不同，结构相同：标记齐全、指标漂亮、检测通过，而学习这件事没有在人身上发生。它们都发生在装置照不到的空转地带。认清这一点，才谈得上下一步——把这两种学生分开。

## 二、分诊：区分"学习真发生"与"只是标记达标"

分诊的目的不是抓人，而是给评价换一个焦点：从"标记齐不齐"换到"过程发生没发生"。方法很朴素——不看结果那一格标记，而是逐条追问过程留下的痕迹。学习真发生过，一定会在人身上留下无法凭空补出的东西；只采集了标记的人，恰恰在这些地方是空的。下面几条问法，可以口头问，可以写进作业设计，也可以做成打分维度。它们共同的逻辑是：真发生过的过程可以被复述、被追溯、被为难；只贴了标记的产出经不起这三件事。

**第一问，能不能复述过程，而不只是复述结果。**让学生说出他是怎么得到这个答案的：先试了什么、哪里错了、怎么改的、为什么最后选这条路。真做过的人有一条带岔路的路径可讲；只采集标记的人只有一个干净的终点，讲不出中间的犹豫和返工。**过程主体性的第一个可见痕迹，就是一条有岔路、有反复的路径。**

**第二问，卡点在哪里，是怎么过去的。**任何真发生的学习都有卡点——某个不懂的地方、某次失败、某个想不通的环节。直接问："这次你最卡的是哪一步？后来靠什么过去的？"答得出具体卡点和具体突破的，过程大概率发生过；答"没什么难的""挺顺的"，要么没真做，要么做的是采集标记而非学习。**没有卡点的学习，通常等于没有发生的学习。**

**第三问，换个条件还成不成立。**把题目的条件改一个、把情境换一个，看他还会不会。真理解的人能迁移，因为他掌握的是过程里长出来的东西；只对齐了标准答案的人，一换条件就塌，因为他采集的是这一道题的标记，不是解决这类问题的能力。**迁移能力是过程发生过最难伪造的证据。**

**第四问，产出里有没有属于他自己的东西。**他的具体例子、他犯过又改正的那个特定错误、他和自己经验接上的那个联系——这些是无法从范文和标准答案里搬来的。通篇正确却毫无个人痕迹的作业，往往正是空转的信号：**它对齐了所有给分点，却没有一处是不可替代的。**

**第五问，为什么是这样，而不是别的样。**让他为自己的选择给理由：为什么用这个方法、为什么下这个判断、为什么排除别的可能。给得出理由的，背后有过权衡；给不出的，多半是照抄或照生成，理由那一层从没在他脑子里发生过。

这五问不需要一次全用。日常抽一两条就能把两种学生分开——关键是把追问变成常态。一旦学生知道评价会追到过程，采集标记的性价比就下降了：因为过程无法临时补，而复述、迁移、给理由这些事，只有真发生过才做得到。分诊本身就是最初的一次纠偏。

### 三、标记装置是怎么空转的

---

要长久地治，得知道它为什么会空转，而不是把空转当成学生品德问题或个别现象。母论文的判断是：空转是这套认证装置的**结构性产物**，只要装置照旧运转，空转就会稳定地、成规模地出现，换一批学生也一样。它由三个互相咬合的子机制推动。

#### 机制一：标记只认得计量终点，认不得过程

装置为了能大规模、可比较地记录，必须把学习压缩成可计量的终点：一个分数、一个等级、一个"通过/不通过"、一个重复率百分比。压缩是它高效的原因，也是它的盲区来源——过程是连续的、个人的、无法直接计量的，一旦被压成终点，就在记录里消失了。系统里因此根本没有"过程是否发生"这个字段，它**只有能力记录终点，没有能力记录终点是怎么来的**。于是两个终点相同的学生，在装置眼里完全一样，哪怕一个真学了、一个只是采集了标记。空转的第一块地基就在这里：装置不是不想看过程，而是它的记录方式从一开始就把过程排除在外了。

#### 机制二：规避标记比真做便宜、安全、可规模化

既然装置只认终点，那么理性的做法就不是去经历过程，而是用最低成本直接产出终点。这就是母论文反复强调的那条不对称：**规避标记远比真做便宜**。真搞懂一门课要几十小时的投入、要忍受卡点、要反复重来；而摸清给分点去对齐、押考点去应试、让基底代写再改到查重通过，成本低一个量级，风险还更小（因为标记合格就安全），而且可以复制给下一门课、下一份作业。成本这么悬殊，采集标记就会成为大多数人的默认策略——不是因为学生坏，而是因为装置给这条路开了绿灯。空转地带里那些"标记合格但学习没发生"的产出，就是这条不对称的量产结果。**只要真做的代价高于规避的代价，空转就会自动扩大**。

#### 机制三：打假与刷标记互相校准、共同升级

学校当然会反应：查重加严、上 AI 检测、监考加密、给分标准细化。但母论文指出，这不是在填盲区，而是在装置内部添一个新组件。新检测一上，规避方也随之升级：降重改写、绕过检测的写法、更精致的应试。检测与规避于是进入一场互相校准的军备竞赛——**打假和造假是同一台装置里两个组件在彼此升级**，谁也没走出装置去碰那个真正的盲区（过程是否发生）。更麻烦的是，每加一层检测，就等于向所有人（包括真学的学生）宣告"我们靠标记来判断"，于是把更多人的注意力从过程进一步吸到标记上，让空转更普遍。升级检测因此往往不是解药，而是空转的加速器。这三个机制咬合在一起，构成一台会自己转、还会越转越快的空转机器。要停下它，不能在装置内部加码，只能换一个方向——让过程重新可见。

## 四、核心重构：让"是否真发生"重新可见的四个杠杆

重构的总原则只有一句：**不要再去升级检测，而要去降低"真做"相对于"规避"的成本，同时提高"规避"被过程性证据识破的概率**。换句话说，把母论文那条"规避比真做便宜"的不对称扭转过来。下面四个杠杆，每一个都配一套能落地的做法。它们不取消原有的标记，而是在标记之外，重新给过程装上可见性。

### 杠杆一：把评价对象从"成品标记"移到"过程痕迹"

只要评价盯着成品，学生就只会去优化成品，过程照样隐形。杠杆是把一部分评分权重，从最终产出转移到过程留下的痕迹上。**落地技术**：要求提交过程档案而非只交成品——草稿的演进版本、卡住时的记录、改错的对照、参考来源的判断说明；作业里明确留出"你最大的卡点和怎么过去的"一栏，并给它计分；口头答辩抽查，让学生当面复述过程。关键在于，这些过程痕迹**无法在最后一刻凭空补出**：真做过的人手边就有，只采集标记的人得反过来重演一遍过程才能补——而那正是我们想要的。这一杠杆直接抬高了规避的成本，因为它把"过程是否发生"变成了要被评的东西。

### 杠杆二：设计"无法靠采集标记通过"的任务

很多任务之所以能被空转，是因为它本身就是标准化的、有标准答案的、可以对齐给分点的。杠杆是重新设计任务，让它只有真经历过程才做得出。**落地技术**：用本地化、个人化、即时性的任务替代通用题——要求结合本班本地的具体数据、要求接上学生自己的某段经历、要求对刚发生的事情当堂回应；用"给出你的判断并说明为什么排除其他可能"替代"选出正确答案"；允许甚至要求学生暴露自己的错误路径并分析它。这类任务的共同点是**没有现成标记可采集**——范文、题库、基底生成的通用文本都对不上这个具体情境，学生除了真做一遍别无捷径。这是从供给侧掐断空转：任务本身不给规避留位置。

### 杠杆三：把 AI 与基底从"规避工具"转成"过程放大器"

AI 和基底让规避成本进一步逼近于零，这是现实，堵不住。杠杆不是禁用，而是改变它在任务里的位置——从"替学生产出成品"变成"逼学生暴露过程"。**落地技术**：要求学生把与 AI 或基底的完整交互过程一并提交，并对其输出做批判——哪里对、哪里错、你据什么改了它；把任务设计成"用基底生成一版，然后指出它三处站不住的地方并改正"，这时评分点落在学生的判断和修改上，而基底生成那部分反而成了被评判的对象；要求学生说明自己在哪一步用了工具、为什么用、它替代了什么、又没能替代什么。这样一来，用基底不再是绕过过程，而是**制造出更多需要学生亲自判断的过程节点**。工具越强，暴露的过程反而越多。

## 杠杆四：用长期追踪替代单点标记

单次的分数、单张的证书都是一个孤立的计量终点，最容易被采集。杠杆是把评价拉长成一条时间线，看同一个人在不同时间、不同情境里是否一致地表现出能力。**落地技术**：建立跨学期的能力追踪，同一项能力在不同任务里反复取样；把"上次的卡点这次过去了没有"作为一个观察项；让学生定期回看并说明自己相比上一阶段变化在哪、根据是什么。单点标记可以靠一次规避拿到，但一条**长期一致的表现曲线无法靠单次规避伪造**——你可以刷过一次考试，很难在半年里、在十几个不同任务里都刚好蒙对。真发生过的学习会累积、会迁移、会在后面的任务里留下前面那次过程的影子；只采集标记的人，每次都要从零再规避一遍，时间一拉长，破绽就露出来。时间是过程最难造假的见证者。这一杠杆把"是否真发生"放进了一个规避成本极高的维度里，也顺带把评价从一次性审判改成了持续陪伴——学生知道你在看的是一条线，而不是一个点，采集单点标记的动机就随之减弱。

四个杠杆合起来，就是把母论文那条不对称反转：让真做因为有过程痕迹可依而变得省力、可信、被承认，让规避因为处处要面对过程性证据而变得更贵、更险、更难规模化。当真做比规避划算，空转自然收缩。

## 五、课堂/学校工具箱

杠杆要落到具体场景里才有用。这一节按场景给出可以直接照做的做法和可以直接照说的话术。原则贯穿始终：不增设新的检测，只增设过程的可见性；不羞辱采集标记的学生，只让真做变得更划算、更被看见。

### 场景 A：布置作业时

**做法**：每份作业在成品之外，固定要求一栏"过程说明"——你卡在哪、试了几种办法、最后为什么选这个、如果用了 AI 或基底用在了哪一步。把这一栏计入分数，占比不必高但必须有。任务尽量绑定本班本地的具体情境，让通用范文和通用生成文本对不上。**话术**："这次作业，做对不是唯一得分点。我更想看到你是怎么想到的、在哪卡住过、怎么过去的——把这些写下来，它和答案一样值分。哪怕你最后没做对，只要过程是你真走过的，我看得见。"这句话直接改写了学生的成本账：过程有分，采集标记就不再稳赚。

### 场景 B：批改与反馈时

**做法**：批改时抽查过程痕迹，对"成品漂亮但过程空白"的作业，不直接判低分，而是追问——请补上你的过程，或当面说一遍。反馈聚焦在过程节点上（"你这一步的判断很关键，说说依据"），而不是只给一个分数标记。**话术**："你的结果没问题，但我从作业里看不出这是你想的过程。你来讲讲你是怎么一步步到这儿的？"这不是质问，是把评价的焦点当面挪到过程上。学生会很快明白：在这个课堂，光有合格标记不够，过程要能被复述。

## 场景 C：考试与考核设计时

**做法：**减少纯标准答案题的比重，增加"给判断并说明理由""指出并改正一处错误""换一个条件重做"这类题。开卷、可用工具的考核里，把题目设计成工具帮不上忙的样子——要个人化、要即时、要迁移。**话术**（对学生说明规则）："这门课的考核，允许你查资料、用工具。因为我要考的不是你能不能找到标准答案，而是你能不能判断、能不能迁移、能不能说清为什么。这些东西，工具替不了你。"把考核从"标记采集竞赛"改成"过程展示"，规避就失去了对象。

## 场景 D：面对 AI 与基底代写的嫌疑时

**做法：**不依赖 AI 检测工具下结论（那只是往装置里再加一个会被规避校准的组件）。改用过程追问：请出示草稿演进、请当面复述、请换个条件即时重做。真做过的人过得去，代写的人过不去，且这个方法不冤枉任何人。**话术：**"我不打算靠检测软件来判断你有没有用工具代写——那种东西我都知道能绕过去。我只请你做一件事：把你写这篇东西的过程给我看看，或者现在当面把核心思路重讲一遍。真是你做的，这很容易。"这话既公道又有效：它绕开了检测军备竞赛，直接回到过程是否发生。

## 场景 E：学校层面的评价设计时

**做法：**在绩点、证书这些单点标记之外，引入跨学期的能力追踪档案，让同一能力被多次、多情境取样；评优、推荐时把过程性证据纳入，而不只看标记堆积；明确向教师传达——学校不以查重率和通过率作为唯一质量指标，鼓励过程性评价。**话术**（对教师团队）："我们要防的不是学生作弊，而是整个评价只剩下标记、没人再看学习有没有发生。别再靠加检测去补，那是无底洞。把力气放到让过程可见上——这是我们能真正拉开差距的地方。"学校这一层若不转，教师个人的努力会被大环境的标记导向不断稀释。

# 六、落地流程

把上面的东西一次性全上，多数老师和学校扛不住，也没必要。给一条可以稳步推进的流程，从一门课、一个班开始，逐步扩到院系。

**第一步，选一个切口。**不要全面铺开。选一门你最有把握的课、一类你最常布置的作业，作为试点。目标单一：在这个切口里，把"过程说明"这一栏加进去并计分。这一步几乎零成本，却已经开始改写学生的成本账。

**第二步，改造任务。**把这门课里最容易被采集标记的那几个任务挑出来，用杠杆二的办法改成"无法靠采集标记通过"的样子——本地化、个人化、要迁移、要给理由。一次改一两个就好，改完观察学生反应。

**第三步，把追问变成常态。**在批改和课堂里，固定使用第二章的分诊问法，尤其是"卡在哪、怎么过去的""换个条件还成不成立"。让学生形成预期：这个课会追到过程。预期一旦形成，规避的性价比就持续下降。

**第四步，安置 AI 与基底。**公开你的规则：不禁用，但要求暴露交互过程、要求批判其输出。用杠杆三的做法，把工具从规避通道改成过程节点。规则讲清楚，比偷偷检测有效得多。

**第五步，拉长时间线。**试点跑过一轮后，开始为同一批学生建立跨任务、跨阶段的能力取样，哪怕只是一张简单的追踪表。让长期一致性进入你的判断。

**第六步，复盘并外扩。**一学期末回看：哪些做法真的把两种学生分开了、哪些流于形式、学生和自己的负担在哪。留下有效的，砍掉变成表演的，再把成形的做法带到下一门课、推荐给同事、上报给院系。**整个流程的判据只有一个：过程是不是比原来更可见了，真做是不是比原来更划算了。**不看新增了多少表格，只看这两件事。

## 七、必须避开的误区

---

这套方法一旦操作走样，很容易退回它想反对的那台装置里去，甚至加重空转。下面四条误区，是实践中最常踩的。

### 误区一：以为加更多标记、更多证书就更可靠

最常见的反射动作，是发现标记不可靠后，去加更多标记：再加一次考试、再要一张证书、再补一层认证。这恰恰是母论文警告的方向——**你没有走出装置，只是把装置加厚了**。每多一个标记，就多一个可以被采集、被规避的终点，空转地带只会更大。可靠性不来自标记的数量，只来自过程的可见性。加标记是在盲区上叠盲区，越叠越厚，越看不见过程。判断一个新举措该不该上，就问一句：它是在增加一个新终点，还是在让过程更可见？前者一律慎重。

### 误区二：把"过程可见"做成表演式档案

要求提交过程档案后，很快会出现一种新的规避：学生不去真做，而是去**补一份好看的过程档案**——事后编造草稿、摆拍卡点、把"过程说明"写成又一篇迎合给分点的作文。这时过程档案本身变成了一个**新标记**，照样可以被采集，空转换了层皮又回来了。防它的办法：过程痕迹要用当场、即时、无法预备的方式抽验——当面复述、当堂即时改条件重做、随机追问细节。表演能准备一份档案，准备不了对任意追问的即时反应。**凡是可以从容准备的过程证据，都可能是新的标记；只有经不起预备的即时证据，才逼得出过程。**

### 误区三：把方法变成惩罚采集标记的学生

如果把过程追问用成抓人、羞辱、扣分的工具，学生只会更防御、更隐蔽地规避，而不会转向真做。母论文的出路是让真做变划算，不是让规避变可耻。**落地上要守住一条**：追问的姿态是"我想看见你的过程"，不是"我要证明你造假"；对过程空白的作业，第一反应是给机会补过程、当面讲，而不是直接定罪。目的始终是把成本账改到真做这边，而不是制造对立。学生一旦感到这是场猫鼠游戏，就会把聪明全用在规避上——那正是空转的燃料。

### 误区四：用升级检测替代重构评价

最隐蔽的误区，是把"让过程可见"又悄悄做回了"升级检测"：上更贵的 AI 检测、更严的监考、更细的查重。它见效快、看起来负责，但母论文说得明白——**这只是给装置内部再添一个会被规避校准的组件，打假和造假会一起升级，盲区一寸没动**。检测可以留作辅助，但它永远回答不了"过程是否发生"这个真问题。凡是把力气主要投在"怎么检测得更准"上的方案，方向就错了；力气应该投在"怎么让真做的过程本身被看见、被承认、被算分"上。前者是军备竞赛，后者才是出路。

## 八、一张随身检查清单

---

把这套方法压成七个问题，随身带着。无论是设计一份作业、改一次考核，还是评估一个学生，问自己这七问，就能大致判断你是在治空转，还是在喂空转。

**一问：我现在评的是标记，还是过程？** 如果只看分数、证书、通过与否、重复率，我评的是计量终点；过程有没有进入我的视野？

**二问：这个任务，能不能靠采集标记通过？** 如果对齐给分点、押考点、用范文或基底生成就能过关，那它天生招空转，需要改造成"必须真做才行"。

**三问：这里，规避和真做哪个更便宜？** 如果规避明显更省事更安全，学生大概率会规避——不是他坏，是我的设计给规避开了绿灯。

**四问：学生能不能复述过程、说出卡点、给出理由？** 做不到，多半是采集了标记而非发生了学习；做得到，过程大概率真在他身上发生过。

**五问：换个条件、换个情境，他还成不成立？** 一换就塌，说明掌握的是这一题的标记；能迁移，说明长出了能力。

**六问：我要求的过程证据，是即时的，还是可以从容准备的？** 可以准备的，会变成新标记被采集；即时的、经不起预备的，才逼得出过程。

**七问：我这一步，是在升级检测，还是在让过程更可见？** 如果答案是升级检测，停下——那是往装置里加码；只有让过程更可见，才是走出空转。

七问背后是同一件事：别再问"标记齐不齐"，改问"过程发生没发生"。母论文把这条出路叫作坚持**过程主体性**——让一次不可替代的学习是否真发生过，重新变得可见、可讲、可迁移、可被承认。检测再精密，也照不进那个盲区；只有把评价的焦点搬回过程，学习才不再是一串空转的标记，而重新落回它本该发生的地方——具体的、这一个学生身上。这不需要更贵的技术，只需要一个决定：不再满足于标记齐全，而坚持去看那件事到底有没有发生。

德麦国际