

生成的非法化：当“还没有成”被体验为“不够格”

作者 黄倩盈 · SDE 学员 · 约 2.7 万字 · 德麦国际 SDE Universes

标准化评价塑造的学习环境里，学习者习得一套比“不敢犯错”更早的自我规制——在念头还没成形为可判对错的表述之前，就自行判定它“不值得继续想”。本文为一间普通教室里一个朴素动作辩护：一个学生悄悄地、不必向谁证明地，把自己走过的弯路认真看一遍。

摘要

本文追问一个被既有教育批判话语长期绕开的问题：在标准化评价长期塑造的学习环境中，学习者如何习得一套比“不敢犯错”更早的自我规制——在念头尚未成形为可被判定对错的表述之前，就自行判定它“不值得继续想”。本文不从“制度压抑个体”的框架出发，而是追踪这套自我规制如何在采分点窄化“正确”、时间与反馈技术架空“不确定中继续待下去”的意愿、以及延迟扣分将“不确定感”条件化为负价信号这三者的交织运作中，被发生出来、被身体化、被体验为“自然而然”的。在此基础上，本文提出全文的核心区分面：“生成性内歧义”——那种在现有概念网络边界处用不成熟的试探去触碰新结构时必然经历的、未被闭合的认知状态——与“技术性失误”在标准化评价的视力范围内共享同一个“未得分”的外观，而这种不可区分性是结构性的，并非评分技术的精度缺陷。本文逐一检讨“过程性评价改良”的尝试，承认其在特定限度内的有效性，但指出其共同边界：评价作为制度物种的构造前提——必须依据已完成的、可被物证的认知产物做出可比较的判断——使得它无法在不扭曲其本质的前提下捕捉“仍在发生中的认知动作”。论文最后重新检视一种微型实践的定位，主张它的价值不在于它是一种“更丰富的评价维度”，而在于它刻意留出这样一个缝隙：在评价制度的外面，让学生有机会自己认领那些曾被他自己标记为“不值得想”的认知经验，在一种不被交出的、低压的书写中，让一种不服务于外部标准的自我关系得以被第一次生成。全文不试图设计一套替代性的新评价制度，而是为一间普通教室里一个朴素的动作——一个学生决定悄悄地、不必向谁证明地，把自己走过的弯路认真地看一遍——给出它应有的严肃辩护。

一、引言：一个比“不敢犯错”更早的动作

在教育评价批判的漫长传统中，有一个几乎是共识性的诊断：以标准化考试为核心的评价体系，使学生“不敢犯错”。这个诊断流传如此之广，以至于它已经变成了一个不需要再被追问的起点——仿佛只要我们能设计出一种“容忍错误”的评价方式，问题就解决了一大半。

本文试图做的事，是在这个共识的起点之前，再往下追问一步。这一步追问的对象，不再是“学生在考场上敢不敢写一个可能错的答案”，而是“学生在考场上、在写作业的桌前、在老师提问的那几秒钟间隙里，他脑子里那个刚冒出来的、还没有变成语言的念头，他自己是怎样对待它的”。

这是一个发生在“犯错”之前的时间窗口。犯错需要的条件是：一个想法已经被表达出来，它有了可以被判定为对或错的形态，然后评价体系给出一个负反馈。这个序列包含三个步骤：生成—表达—被否定。既有批判的主要火力集中在第三步（被否定），部分延伸到第二步（表达的风险），但几乎没有触碰到第一步——那个“生成”动作本身，在什么条件下，会被学习者自己提前终止。

本文的核心主张是：在长期受可比较性压力塑造的学习环境中，学习者不仅学会了不在试卷上写下“不保险”的答案，而且学会了一个更早、更隐蔽也更致命的自律——当他脑子里出现一个模糊的、不确定的、尚未形成清晰表述的认知冲动时，他不再把它当一回事。他不是“不敢说”，他是“不觉得值得想”。那个念头在刚刚冒出来的几秒钟里，还没等到他决定“要不要写出来”，就已经被他自己归入了“瞎想”的类别，然后被妥帖地翻篇了。

我把这种发生在认知生成阶段的自我弃守，称为“生成的非法化”。这个命名需要被严格限定：它不是指“某个想法被外部权威判定为不合法”，而是指“学习者自己已经把生成那些尚未成形的念头这件事，经验为一件不值得认真对待的事”。这其中蕴含着一种权力运作方式的深层迁移——当权力需要禁止一个行为时，它至少承认那个行为是存在的、是有可能发生的；但当权力运作到可以让那个行为在发生之前就被当事人自己判定为“不值得发生”时，权力就不再需要以禁止的面目出现了。它变成了当事人自己的理性选择。

这一命名必须与两个最容易混淆的概念划清边界。第一个是“习得性无助”（learned helplessness）。Seligman（1975）及其合作者通过系列经典实验确立的这一概念，其形成前提是反复的失败经验和对这些经验的挫败性归因——学生因为长期学业失败，最终“觉得自己不行”，放弃努力。而“生成的非法化”所描述的现象，不需要以失败为前提。一个考试成绩相当优秀的学生，同样可能已经内化了这套自我筛查语法：他不是“不行”，而是已经把“行”的标准从“能理解、能推演、能生成新解法”，置换成了“能稳定复现标准答案”。在这个置换完成的瞬间，他的生成能力依然存在，但它的合法使用场景被收缩到只剩下那些“明确有分”的赛道。这里的放弃，不是放弃努力，而是放弃对那些“在现行评价体系中沒有明确回报的认知活动”的认真投入。它不一定伴随挫败感，更多的时候，它只是一阵轻微的“算了不想了”的释然。

第二个是“自我审查”（self-censorship）。自我审查通常预设一个明确的外部压力源和被主体感知到的禁止——一个记者知道某条报道不能写，一个公民知道某句话不能公开说。这种约束是策略性的、有意识的、针对特定内容的。而“生成的非法化”所处的位置比这更早。它不是“这个想法不能说”，而是“这个想法还没成形、我还不确定它是什么——那就别费劲去把它成形出来了”。它运作在沒有明文禁令的地带——没有任何一条校规写着“不许用自己的话理解问题”——它依赖的恰恰是这个“沒有禁令”的模糊性。学生需要的不是压制一个已经成形的表达冲动，而是训练自己的注意力不再往那些“不成形的东西”上流。

与这两个概念的严格区分，是本文核心命名得以成立的初步条件。但仅做这一步澄清还远远不够——更根本的问号悬在它的头顶：即便“生成的非法化”与习得性无助和自我审查都不同，它是否仍然只是用一个新名称重新描述了某个已有的理论传统早已处理过的问题？

这个追问必须被正面回答。本文的回答分两层。

第一层，与教育社会学中“隐性课程”（hidden curriculum）传统的区分。自Jackson（1968）在《教室中的生活》中提出这一概念以来，半个多世纪的研究已经充分论证了：学校通过制度安排、空间组织、师生互动、奖惩结构等，向学生传递着一套未被明言的规范——耐心等待、服从权威、与同辈竞争、接受外部评价的合法性。学生学会“看老师眼色行事”，学会“找出老师真正想要的那个答案”，这些观察都是隐性课程研究的经典成果。“生成的非法化”与隐性课程存在明显的交集：它处理的同样是“未被明言但被制度性地传递的规训”。但二者的区分面在于时间尺度。隐性课程研究主要揭示的是“应当做什么”的价值与行为规训——什么样的表现会被奖励，什么样的态度会被惩罚，学生据此调整自己的行为策略。而“生成的非法化”捕捉的是一个更早的、发生在认知流程内部的时间节点：不是在行为应当如何表现的意义上调整自己，而是在“这个念头本身值不值得继续往下想”的层面，就提前终止了认知过程。隐性课程告诉你“在课堂上要举手才能说话”；生成的非法化发生在你举手之前——你甚至不觉得那个刚冒出来的模糊念头值得被整理成一个可以举手说出口的问题。这不是行为规训，这是认知生成过程的合法性被预先撤销。Jackson和后来的研究者确实观察到了学生对“等待线索”（waiting for cues）的策略性顺应——他们学会了从教师的提问方式中研判什么答案最可能被接受。但他们没有追问：那种在没有任何外部线索可资研判时、从自己内部冒出的、还不确定走向何方的认知冲

动，学生自己是怎么对待它的？而本文追踪的，正是这种“自己对待自己的模糊念头”的方式。

第二层，与教育评价社会学中“反拨效应”（washback）与“评价退化”论述的对接。近四十年来，Crooks（1988）、Gipps（1994）、Broadfoot（1996）、Torrance（2007）等人已经系统论证了一个判断：高风险评价不仅塑造“教什么”，更深刻地塑造了“如何学”与“如何认知自己的学习”。尤其是Torrance关于评价退化的分析，已经触及了过程性评价在被制度化后如何蜕变为新的表演形式——当“过程”本身被纳入评价，学生很快学会如何展示一种“看起来像认真探索的过程”。本文在这一判断上直接继承了上述传统的地基。但本文在此地基上追问的是一个上述研究尚未充分展开的问题：当评价制度长期稳定运作，学习者对“什么算有效证据”、“什么算值得投入时间”的认知，是否可能不再依赖于对评价规则的有意识研判，而是沉淀为一套自动化的、前反思的认知倾向——一套不被当事人自己识别为策略的策略？上述文献主要描述了学习者在策略意识层面的适应——他们知道自己在“应对考试”，知道自己在“把握评分标准”。而本文所追踪的“生成的非法化”，它的特异性恰恰在于：学生在放弃那些模糊念头时，他并不觉得自己正在“应对”什么。他只是觉得那个念头本身就不值得想。他不需要“揣摩评分标准”——他的注意力已经被训练得不会再往那个方向流了。这不是策略，这是策略被自然化之后剩下的东西。它要求我们把分析单元从“个体对评价规则的有意识适应”下沉到“评价规则在日常教学事件中如何无声地重塑个体的认知倾向本身”。

这一下沉动作，也划定了本文与Dweck（2006）“固定型思维”（fixed mindset）及Steele与Aronson（1995）“刻板印象威胁”（stereotype threat）这两条另一重要研究脉络的区分面。固定型思维的核心机制是：个体将自己遭遇的认知困难归因于“我能力不足”这一实体化信念，从而回避挑战。刻板印象威胁的机制是：个体因担忧自己的表现会印证所属群体的负面刻板印象，而产生焦虑、认知资源耗竭与表现下降。这两种机制的共同前提是：困难已经被个体感知为一种威胁——要么威胁到自我能力评价，要么威胁到群体身份认同。而“生成的非法化”发生在比这更早的窗口：那些模糊念头在刚刚冒出来、还没有被主体评估为“困难”还是“机会”的时候，就已经被无意地搁置了。它不是“我害怕做错所以不做”，而是“我根本不觉得这东西值得被当作一件事来做”。这不是动机层面的自我保护，而是注意层面的提前撤离。固定型思维与刻板印象威胁都预设了主体已经对他的认知处境做出了某种评价；而生成的非法化要追问的正是：在什么制度条件下，主体会不再对某些认知处境做出评价——因为那个处境本身，在他的经验世界里，已经不再以“需要被认真对待的选项”的形态出现？

与上述三条脉络的逐一对质，是本文核心命名得以成立的硬条件。如果“生成的非法化”可以被washback、固定型思维或隐性课程中的任何一方充分覆盖，它就是一个多余的命名。本文的判断是：它切开了它们各自装不下的一个区分面——一个比行为策略更早、比动机归因更早、比隐性课程的价值规训更微观的时间窗口。这个窗口内的经验，不是“对评价规则的策略性适应”，不是“因能力信念而产生的回避”，而是认知注意在制度条件的长期渗透下发生的结构性重组。

这一重组的具体发生机制，将在第二节中展开追踪。但在进入机制分析之前，有必要先锚定一个时间节点。这关系到一个看似平凡、实则关键的问题：为什么是“念头”？

一个被忽略的时间节点：从感觉到命名之间

任何有过学习经验的人都能回忆起这样一种时刻：面对一个问题，你隐约觉得自己抓住了什么，但你还说不出来那是什么。它还不是一个可以被写进答题框的句子，也不是一条干净的、可以在心里默念的思路。它更像是一种介于身体感受和语言之间的认知信号——同时包含着方向感（“好像和那个东西有关”）和不确定感（“但我还说不清到底是什么关系”）。你在草稿纸上画了两笔，画得不成形状；你在心里默推了两步，推得磕磕绊绊。你可能在第三笔就放弃了，也可能继续了十分钟，但最终因为“实在想不出”而搁下。让这个时间片段在教育上值得被严肃对待的，不是它是否最终产出了一个正确答案，而是它是否在某种制度条件下被过早地终止了——而且终止得如此自然，以至于当事人自己都不觉得有什么东西被终止了。

在本文引入一个专门的概念来捕捉它之前，必须先精确描述它的经验品质：它不是能力不足导致的错误，也不是动机不足导致的放弃。它是一种“半生不熟”的认知摸索——旧有的解题模式套不上去，新的结构还没有成形，学习者在这个缺口处用自己的不成熟的、试探性的认知动作去架设连接。它的推进是摸索性的，不是算法性的；它的步伐是不连续的，不是渐进式的；它的失败代价无法提前估算，因为它还不确定自己会在哪里失败。这种状态对深层学习的认知价值，已经有大量文献从不同角度予以论证——皮亚杰的“认知失衡”、维果茨基的“最近发展区”、Kapur的“生产性失败”——这些构成了本文的重要对话资源，但都不是本文的起点。本文的起点是：如果这种状态与真正的能力不足在制度视野中不可区分，那么学习者需要多长时间，才能学会不再把自己扔进这种状态里去？

二、窄化、不划算化与条件化：生成的非法化如何在日常教学中被交织运作出来

“生成的非法化”不是某一个特定制度条文的结果，也不等同于教师的一次不当反馈。它是在三套日常教学运作的相互渗透、彼此喂养中，被逐步沉淀为学习者的认知习惯的。这三套运作在分析上可以被分别追踪，但在经验层面，它们在同一间教室、同一节习题课、同一道作业题中交织出现，彼此加固。要理解“生成的非法化”如何发生，不能把它们拆成三个独立的原因，也不能把它们罗列为一个前后相继的因果链：先有采分点窄化，再产生不划算感，最后沉淀为筛查语法。这种递进框架的缺陷在于：它把复杂的同时互生简化成了单线进程，它把“延迟扣分”仅仅当作第三层的一个节点，而忽略了它恰恰是连通三层的那根铁索。

在本节中，我将选择一个具有时空厚度的事件切片——一次课后作业从下发、到学生在家完成、到次日发回的完整周期——来展示三套运作如何在这个周期内同时发生、互相喂养，像三条绞在一起的绳股那样，一道一道地绞紧。

一次作业周期：一个复合事件

假设某初中数学老师在周五下课布置了一套习题。她是一位负责任、在大多数学生眼里“讲得还不错”的老师。这套习题的内容是上周教的一元一次不等式的应用题，共十道，其中前八道都是标准题型，直接在已知情境中设未知数、列不等式、求解；最后两道是略有变式的情境——比如生活情境，不是从课本上的标准应用题模板改编的，而是需要学生自己从题干的描述里识别出哪个量是未知数、用不等式把一组限制关系翻译出来。

老师在布置作业时说的原话是：“最后两道题可能有点绕，但不要怕，自己先想一想，想不出来就空着，下周一我们讲。”她说这句话的时候是真心的。她希望学生自己想一想。

让我们先悬置“学生”这个抽象的集合名词，进入一个具体的学生。叫他林。他在班里成绩中等偏上，不算顶尖，但对数学不排斥，也没到热爱的程度。他的草稿纸平时写什么并不被检查——这不是一项有制度分量的材料。他在这个周末写了前八道题，很顺。第九道题卡住了。他读了两遍题干，在草稿纸上画了个大概的数量关系图，试着设了一个未知数，列了一个不等式，发现解出来不对。他擦了，重新读了一遍题，想从另一个条件切入。他感觉这道题“和课上讲的不太一样”，但他说不清哪里不一样。他在这个状态里待了不到五分钟，然后他在心里做了一个极快的、但当时他自己未必有意识地分析的决定：先空着，做第十道。第十道也是变式，他试了一个思路，没走通。他看了一遍题目，再次感觉到那种“和课上讲的不太一样”的模糊不适。这一次他没有再试。他合上作业本。晚上他妈妈问作业做完没，他说做完了。这不是撒谎。在他自己的界定里，那两道题他“想过了”但“不会”——用他的话对自己说，就是“脑子转不过弯”。周末

剩下的时间，他没有再打开作业本。周一早上，他在学校旁边等同学的几分钟里用搜题软件扫了这两道题，看了标准解法，心里“哦”了一下，抄了上去。上课时老师讲评，他跟着听，发现思路其实不复杂，觉得自己“如果再多想一会儿大概也能想出来”。但那个“再多想一会儿”的片刻，在周末的某一个晚上，已经不留痕迹地过去了。

这个周期看起来平淡无奇。它没有任何戏剧性的对抗、没有任何教师对学生的否定性评价。事实上，老师甚至还主动鼓励了“想一想”。但它完整地包含了生成的非法化得以发生的全部要素。让我们回到这个周期中，逐一检视那些要素在同一个事件中如何同时运作。

窄化：“正确”从“理解通达”滑移到“表述吻合”——但它不是第一层，它是持续的背景辐射

首先需要指出一件初看反直觉的事：林在第九题面前的卡住，并不是采分点窄化造成的。他没有参考任何评分标准，没有查阅历年真题的答案格式，甚至没有去想“标准答案长什么样”——他是在自己家里、不受时间严格限制地、完全与任何外部评价隔离地做题。那么，采分点窄化在哪里起作用？它不在这一道题、这一次卡住的时刻起作用。它起作用的时刻，在此前更长的时间里。

在此前的无数次作业、测验、课堂提问中，林已经反复经历了一种极轻微的、累积性的经验：他自己用自己的语言组织出来的理解，与课本上的表述、与老师黑板上的标准解题步骤、与标准答案里的措辞之间，持续存在着一个几厘米的落差。这个落差有时被扣分，有时不被扣分。老师未必每次都严格地把“用自己的话表达正确理解”判定为不完整，但可比较性的底层压力要求她对措辞差异保持一种持续的低度敏感——因为如果她对措辞完全放手，不同学生给出的含混表述在评分时就无法被一致处理。这个矛盾不是任何教师的个人选择造成的。只要评分需要在不同评卷人之间取得可接受的一致性，采分点制度就必然倾向于把“正确”操作化为“可以被明确标识的表述单元”——关键词、关键步骤、关键结论。这不是考试制度“堕落”了才会发生的事。这是评价信度的技术要求在制度层面产生的一个非意图后果：为了保障“谁改都一样”，制度必须把“对的理解”翻译为“对的写法”。

这个翻译一旦在课堂日常中被反复兑现，它在学习者经验侧产生的效果，就不是任何一句“用自己的话答对也算对”能够抵消的。因为学生在一次重要考试中因为措辞不规范而被实际扣分时的那种切肤经验，与老师在讲评时说的那句口头声明，在身体记忆中的分量是不对等的。更关键的是，这种翻译不是在一个单独的时间点上发生的——它是一种持续弥散在每一次作业批改、每一次课堂提问、每一次“某某同学的回答基本对，但表述还是要再规范一些”的日常反馈中的背景辐射。学生不需要某一次被严厉否定，他只需要在足够多次的日常反馈中捕捉到一个稳定的模式：写得越像标准答案，越安全；自己组织语言，永远有一点点悬。这种模式一旦被身体捕捉到，它就不再需要每一次都重新被认知系统处理。它变成了一个在认知启动之前就已经生效的注意力过滤器——不是“我要想一个标准答案出来”，而是“我不去想那些我暂时想不出标准表述的东西”。林在周末面对第九题时，那个过滤器已经在运转了。他没有去想“标准答案是什么”——他只是觉得这道题“和课上讲的不太一样”。这个“不太一样”的感知本身，就是一种已经被窄化过的认知信号。它的认知前提是：他已经默认，值得认真对待的问题，应该长得像课上讲过的那些问题。

不划算化：时间与反馈技术如何让“扛住不确定”变成一件需要额外意志力的事

现在让我们转向这个作业周期的第二个断面。林在第九题面前卡了不到五分钟——没有时间的严格截止，也没有人在旁边催他。那么，支撑“不确定性不划算化”的那个时间压力，在哪？

它不在这一次作业的完成时限里。它在此前更长的时间里，被一件事反复地夯进了他的经验结构：在一堂四十分钟的课里、在一个两小时的考试中、在从放学到睡觉之间有限的时间里，遇到卡住的题，先跳过去——这是一种被所有应试策略反复肯定的高效做法。从功利角度看，这个策略完全合理。但它在学习者的经验结构里留下了一个持久的副作用：它把“继续想下去”在默认状态下标记为“低优先级”。那些需要长时间浸泡、反复试错的认知过程，被挤到了时间表的边缘，只能在“正事做完了还有空”的时候才被允许发生——而大多数时候，“正事”永远做不完。当林在周日晚上面对一道一时想不出思路的变式题时，他身体里那个已经被上千次时间管理决策训练过的本能，比他更早地做出了判断：这件事的优先级，排在其他还没做的作业、还没复习的课文、以及“早点睡明天要早起”的后面。他不是有意识地盘算时间——他只是感觉到了那种熟悉的“算了先空着”的释然。那阵释然来的时候，不太像焦虑的缓解，更像一种认知肌肉的自动卸力。

但光是时间压力，还不足以让“扛住不确定”变成一件需要额外意志力的事。另一件事与它同时发生：搜题软件的随手可得。

这里需要非常精确地理解这个问题，否则极易滑入一个被反复提起的稻草人式回应：搜题软件“只是提供了选项”，它并没有强制学生不准自己想。学生完全可以先自己想二十分钟，再搜答案。确实，“先想后搜”和“直接搜”都是可选动作。但必须被追问的是：这件事的成本结构是什么？“直接搜”能在几十秒内解除认知不适——看完答案，心里“哦”一下，那阵模糊的、说不清的不舒服就过去了。而“自己先扛二十分钟”的成本是什么？持续的不确定、反复的试错、到最后可能还是没想出来——而且想不出来这件事，是没有任何制度回馈的。作业本上空着的地方第二天老师可能只是匆匆扫一眼，学生的名次排名、评优加分、家长的满意程度，都取决于最终呈现的正确率，而不是草稿纸上划掉的那几条死路。

当方便的成本几乎为零、扛住的成本是持续的认知不适加上不确定的回报时，“自己先扛”就不再是一种默认的、被环境自然而然承接的选择，而变成了一种需要每次重新说服自己的个人修行。“扛住不确定”从一项被制度性地镶嵌在学习过程必经路径中的默认能力训练，变成了一项需要持续内部斗争的奢侈。

林在周一早上用搜题软件扫那两道题的时候，他刚在冷风里等同学等了五分钟，手机就在口袋里。他没有盘算“要不要先自己再想一想”。他用那个动作结束了一件他其实周六晚上就已经决定不再继续的事。搜题只是让这个结束来得更干净。

条件化：延迟扣分如何把“不确定感”本身变成一个负价信号

现在我们可以进入这个作业周期中最关键、也最隐蔽的那个断面。很多人会问：如果林并没有因为“自己想了想”而遭到任何惩罚——他空着的题，老师没批评他，他甚至最后抄上了正确答案——那么，“生成性内歧义”被注销的证据在哪？他没有遭受失败，也没有被扣分。他的自我筛查感从何而来？

答案不在这一次作业里。答案在更早的、已经被他忘掉的无数次作业里。

我把这种经验称为“延迟扣分”：学生在提交一份作业、一张试卷时，他并不确定自己写的某个解法是否正确，但他还是写了。在他写的那一刻，他的内部经验是“好像可以这么想，但不确定”。然后作业发回来，扣分了。这个扣分发生的时刻，与他自己当初感到“不确定”的那个时刻之间，有一个时间间隔——可能是一天，可能是一周。在间隔期间，那道题的内部记忆已经流失了。他不太能回忆起当初那种“不确定”时在脑子里具体跑过了哪些试探性思路，但是他的身体记得一件事：那种不确定的感觉，上一次出现之后，跟过来的是扣分。

下一次，当类似的不确定感又冒出来时，扣分的记忆比试探的思路更先到达。他的身体先于他的推理做出反应——一阵轻微的烦躁、一种“算了”的释然、一股把草稿纸翻到下一页的冲动。他不一定在意识层面形成过“这种不确定大概率会导致扣分，因此不值得继续想”的判断。他的认知系统已经自动把那种“不确定

感”和“不好的结果”之间的那个时间间隔压缩为同时发生。他不再需要等到扣分出来才判定这个念头不值得。那种不确定感本身，就已经是一种不舒服的身体信号。回避这个信号，就是回避不舒服。这不是推理，这是身体被条件化之后的自动反应。

这一机制经常被心理学传统描述为“经典条件作用”的类比：中性刺激（不确定感）反复与厌恶刺激（扣分反馈）在时间上配对，最终中性刺激自身就触发了回避反应。但这只是一个粗糙的近似。本文并非在主张学生真的像巴甫洛夫的狗那样，被简单刺激-反应联结所驱动。人类学习者在面对不确定感时的回避，远比条件反射复杂。它不是一个人造铃声引发唾液分泌——它卷入了一个更精密的多层认知加工链：不确定感被元认知系统标记为“某些东西不对”的信号，这个信号激活了关于自己当前知识状态的评估（“这道题我可能没掌握”），评估又触发了对后续风险的下意识预判（“如果继续花时间，大概率做不出来并且浪费了做其他事的时间”），而延迟扣分的累积性记忆则不断地给这条加工链灌注情感色调——让这个加工链不再是一个中性的、冷静的风险评估，而变成了一个被负价情绪包裹的、从身体层面就想回避的经验流。这里的关键，不是简单的S-R联结，而是“不确定感”作为一种认知信号，在长期被与延迟扣分配对之后，它的情感色调从“中性-待探索”被逐渐转向了“负价-应回避”。学生在回避它时，他的元认知系统甚至可能在帮他快速做出一个隐性判断——“根据以往的经验，继续在这上面花时间的期望收益是负的”——这个判断本身是理性的，但理性所依托的那个“以往经验”的数据集，本身就已经被评价制度的窄化效应所污染了。

这正是本文需要揭开的那个制度-认知耦合的现场：不是学生做出了错误的成本收益分析，而是他用来做成本收益分析的那个经验库，已经被一套只奖励“表述吻合”而看不见“认知试探”的制度长年累月地灌浆加固了。他的理性是基于一个被扭曲的数据集做出的——而他的身体，比他的理性更早地适应了这个数据集。

至此，林的故事不再是一个孤立的案例，而是三层交织的一个具体凝结。他在周末面对那两道变式题时，采分点窄化作为背景辐射，已经让他对“和课上讲的不太一样”的题目带有一种无法言明的距离感（窄化）；他平常被时间和即时反馈技术反复塑形的注意力分配倾向，让他在卡住的第五分钟时自动地将“继续想”降级为低优先级（不划算化）；而他此前无数次的作业中经历“不确定→扣分”的累积配对经验，让他的身体在“不确定感”刚一冒出时就微调了认知姿态——不是推理性地决定“放弃”，而是注意力自然地不再往那个方向继续流（条件化）。这三者不是三个分开的步骤。它们是同一个体验事件的三股绳。

不是抑制，是雷达关掉了

有一个重要的澄清必须在此做出。生成的非法化不是压抑。压抑意味着有一个明确的、可以被意识到的东西被按了下去——学生在压抑时，他知道自己正在压住一个想法。而林在合上作业本的那一刻，他并不觉得自己压住了什么有价值的思维。那个念头在他合上作业本时，严格来说还没有机会发育成一个念头。它只是一点模糊的、介于身体感受和语言之间的认知涟漪。他甚至不记得自己有过这个涟漪。他只觉得“最后两道题不会做”是一个不需要再被深究的事实。

这不是内心的一个批判声音对他说“你这个想法不值钱”。这是他的注意力雷达在那些不成形的东西出现之前，就已经被调到另一个频道上去了——那个频道的预设过滤条件是：不清晰的信号，不接收。雷达不是被关掉了，是被重新定义了“什么算信号”。而他的整副认知身体，在无数个作业周期中，已经精熟于这套过滤——快到他意识不到过滤在发生。

三、生成性内歧义：一块在评价视力范围内不可见，但在认知发生上不可替代的飞地

如果上一节的论证成立——学习者确实习得了一套在念头成形前就自动筛查的认知倾向——那么接下来的追问必然是：被这套倾向筛掉的东西里，有哪些是即便在教育价值尺度上也值得被严肃保护的？这个问题不能笼统地回答“创造力”“想象力”“批判性思维”。这些大词太容易让讨论滑入一致赞美的安全区，而错过了那个真正需要被打开的区分面。

本文在此做出的辨别是：在那些被提早筛掉的模糊念头中，有一类具有独特的认知发生学身份——它不是胡思乱想，不是注意涣散，不是能力不足，而是一个人正在用自己的现有概念网络去触碰一个新的、他尚未掌握的结构时所必然经历的那种“半生不熟”的认知状态。本文称其为“生成性内歧义”。

需要在此先明确两个核心概念的分工。本文用“生成性内歧义”来指称被“生成的非法化”机制所过早筛掉的那些模糊念头中、具有独特认知价值的那一类。“生成的非法化”是制度-认知耦合下形成的自我筛查倾向，是机制；“生成性内歧义”是这种机制所注销的认知状态中最值得被严肃对待的那个子集，是内容。前者回答“什么东西在筛”，后者回答“被筛掉的东西里，哪些是丢了之后真正可惜的”。本节将展开后者的认知身份与制度困境。

什么算生成性内歧义

生成性内歧义不是一个可以被已有心理学概念1:1替换的命名。它与“认知失衡”（cognitive disequilibrium）和“生产性失败”（productive failure）既有深层关联，又有不可化简的区分面。这三者之间的关系不是对立，而是叠加：生成性内歧义是认知失衡的一个特殊亚型，是生产性失败的一个被制度条件压缩过的变体。但正是在这种“特殊”和“被压缩”之中，它切出了一个既有概念装不下的区分面。

皮亚杰（1936/1952）的“认知失衡”描述了认知系统在遭遇无法被现有图式同化的新异信息时产生的内部冲突。这是认知发展的驱动力。但认知失衡是一个覆盖范围很广的概念：它可以发生在幼儿发现敲击积木的不同部位会发出不同声音的时刻，也可以发生在物理学家发现现有方程无法解释新实验数据的时刻。生成性内歧义共享了认知失衡的核心驱动力——旧结构不够用，新结构尚未形成——但它在以下三个特征上同时成立，而这些特征在认知失衡的一般描述中并不总是同时出现。

第一，它发生在学习者现有解题模式不足以覆盖一个新问题结构，但他尚未正式学习那个可以解决它的新概念的时刻。他不是“检索已知知识失败了”，而是在“用已知知识延伸出去的触角已经开始碰到那个他还不能命名的东西”。一个初中生尚未学习牛顿第二定律，但在面对一个非匀速运动物体时，他的脑子里同时出现了几条试探性思路：“加速度是不是跟合力成正比？”“是不是跟质量有关系？”“如果合力不变，加速度会变吗？”这几条思路彼此不完全兼容，他每一条都说不周全，但他同时被它们拽着。他不是迷失了——他的认知系统正在一个还未被正式概念装备的地带里尽力找路。这与能通过检索公式来解决的困境不同：如果他已经学过 $F=ma$ ，这道题就只是一个知识调用效率问题，不存在内歧义。这是生成性内歧义与认知失衡的第一次分化：它特指那种发生在“尚未被正式教学覆盖、但在旧有概念网络边缘已经可以被隐约感知”的地带里的失衡。

第二，这种内歧义不能被“检索已知信息”消解，但也不是全无方向的盲目摸索。它与Kapur（2008, 2012; Kapur & Bielaczyc, 2012）的“生产性失败”（productive failure）在这个点上存在最深层的交集。Kapur团队通过大量实验论证了一个判断：让学生在缺乏先备知识的情况下进行生成性探索——即便伴随高错误率——比先直接讲授再应用更能促进深层学习与迁移。这个判断本身，为本文所辩护的生成性内歧义的认知价值提供了强有力的实证背书。但本文必须指出的是，Kapur的实验情境与本文所描述的日常课堂情境之间，存在一个经常被忽略的制度条件差异。在生产性失败的实验设计中，学生所处的探索情境是制度性地免于外部比较后果的：他们知道这次探索不打分、不排名、不影响任何评价结果。换言之，实验情境本身人为地保护了探索行为的合法性。而本文所追问

的恰恰是：在真实课堂中，这种合法性被评价制度的日常运作持续地消解掉了。学生不是在“免于评价后果”的庇护所里进行探索，而是在同一次作业、同一张试卷上——在“这里会扣分”和“这里算探索”之间没有隔离带的连续空间里——被反复告知：做错题和做不出题，在外观上共享同一个未得分。Kapur证明了这种探索在被保护的时候具有极高的学习价值；本文追问的是，在真实的评价制度中它究竟能不能被保护住，如果保护不住了，学生又会学到什么。这两条追问是互补的，不是互相否定的。本文不是在与Kapur争夺“谁发现了探索的价值”——本文接过了这个判断，然后把它放进了一个更棘手的制度情境中，去观察那个制度情境如何把这种探索的被制度承认的可能性榨干。这是生成性内歧义区别于productive failure的不可化简面：它不是“有指导的探索活动”，而是“那种活不过评价制度的生存筛选的探索”。它不是在实验中被设计出来的学习策略，而是在教室日常中被评价制度持续压迫、因而趋向于灭亡的认知本能。

第三，它的推进方式是试探性的，不是算法性的。学习者在面对生成性内歧义时，他的主要认知动作不是“排除错误选项”，而是在几团模糊的、彼此竞争的半成型路径之间来回踱步——试试这条，觉得不太对，退回来重新框定问题；再试试那条，觉得好像气息顺了一点，但还不稳。他可能在这个状态里待很久，而从这个时间段的外部看，他是“卡住了”。他不是在向正确答案稳步逼近，而是在看不清路的情况下自己摸索。这一特征使它与“刻意困难”（desirable difficulties, Bjork, 1994）形成了部分重叠，但同样不可化简。刻意困难的经典范式——分散练习、测试效应、交错练习——都依赖于已知内容的可控参数调整：你把已知的学习材料分散到更长的时间轴上（而非集中在一次学习），制造检索难度，从而增强长期保持。它的困难是“已经被已知结构框定了困难参数的困难”。而生成性内歧义恰恰处在那种困难参数无法被事先设定的地带——学习者不知道他正在试探的这几条路里，哪条会把他带到一个可以与已知结构相衔接的地方。他面对的不是“我知道这个材料，但需要花力气检索它”，而是“我不知道这个问题的结构是什么，我在试探中试图让它的结构变得可见”。一个是已知内容的存取难度，一个是未知结构的摸索难度。二者在认知品质上有质的区别。

这三个特征合在一起，勾勒出了生成性内歧义在认知发生学中的特殊地位：它不是认知系统运行中出现的噪声，不是需要被清除的混乱。它是认知结构即将发生重组的那一会儿——旧结构已经不够用，新结构还没有形成，学习者在那个缺口处用自己的不成熟的、试探性的认知动作去架设连接。这一会儿，是任何深层理解都无法跳过的。

它与“技术性失误”的制度性不可区分

生成性内歧义在教育中最容易被误认的，是那种因为公式记混、审题疏漏、计算粗心而产生的技术性失误。一个学生做错了一道几何证明。他可能压根没理解“辅助线”的构造逻辑，只是胡乱画了几笔，错了；也可能他已经隐约感觉到图形里有一种对称性，他试图画一条从没见过的辅助线来捕捉那种对称，试着写下几条并不规范的推导步骤——它们在外观上不整齐，术语不一定准确，中间甚至有一个逻辑跳跃——但他正在自己摸索一条老师没教过的路。这两种“错”，在评分时都是未得分。在大多数课堂讲评中，它们都会被归入同一类操作——“这道题做错了，我们来看正确做法”。没有任何制度化的机制去区分二者。

这种不可区分性不是由于教师的疏忽或评分制度的设计不够精细。它是结构性的。要理解这一点，必须退后一步去看“评价”这个制度物种的认知前提。任何有分配后果的评价，都必须能够回答一个合法性问题：“凭什么你给他分不给我分？”为回答这个问题，评价必须依据能够被第三方复核的证据——不能是评卷人的印象，不能是学生自己的声称，必须是写在答卷上的、可以被标画、可以被核对、可以在申诉时被重新翻阅的已完成且固定的认知产出物。我们可以把这种证据方式称为评价制度的“物证前提”。

这意味着，在评价制度的条件下，一个认知事件要获得“被承认”的资格，它必须能够被转化为一种可存档的物质形态。未完成的思路不行——它不在答卷上。草稿纸上划掉的线条不行——它不具有被复核的稳定形态。内歧义被体验却未被命名的那一刻的身体状态——更加无法进入。物证前提如此深刻地被嵌入教育评价的一切常规操作中，以至于我们常常忘记它根本不是“自然的”——忘掉它是一项要求，也是一项排除。

生成性内歧义恰恰在这一点上与评价的构造前提正面相撞。它是一种正在发生、正在流动、正在成形过程中尚未凝定的认知动作。它的价值，不在于它最终产出了什么可以被打分的“正确结果”，而在于它在那段不确定的摸索中，锻炼了学习者那种“在没有人告诉你往哪走的时候自己摸索”的认知能力。这个锻炼的过程，是无法被物证化的——你可以事后追认它，复述它，写下它的叙事，但你无法在它还在发生的时候把它直接呈交为一组可以被评分的证据。而它与技术性失误在“未得分”这个最终呈现上是完全相同的——卷面上的物证材料只能告诉学习者“你答错了”，但无法告诉他“你是在试探中走偏了”还是“根本不知道该做什么”。

学生很快就能从累积经验中把这一切算清楚。他不需要老师刻意打击他——他只需要观察到“认真试探和做错在成绩单上呈现为一样的结果”，就足以得出结论：试探，从制度理性的角度看，不是一种有竞争力的学习策略。于是，第二节所描述的那套自我筛查倾向，在这里达到了它最锋利的应用情境。生成性内歧义——那个“好像快摸到了但还说不太清”的状态——在体验上天然就带有“不靠谱”的外观：不整洁、不高效、不确定、答不对。那套内化的倾向几乎就是为它量身定做的反应模式：“你还不清楚——说明它不靠谱——别在这种事情上浪费时间。”此时此刻，放弃发生得完全没有对抗、没有禁令、没有可以被申诉的不公正。它只是理性选择。

四、尺子的构造前提：为什么改良无法从内部完成

在前述分析的地基上，本节转向一个更棘手的问题。如果生成性内歧义在既有评价体系中遭遇的是结构性的不可区分——而不是偶然的疏忽——那么大量以“用更细致的评价维度去捕捉它”为宗旨的改良方案，究竟能在什么程度上解决这个问题？本节将逐一检视两种最有影响力的改良进路：档案袋评价与评价量表。它们的善意毋庸置疑，但它们各自在遭遇“生成性内歧义”时，都撞上了一面由评价制度的构造前提筑成的墙。

两段改良简史：档案袋与量表

档案袋评价（portfolio assessment）兴起于二十世纪八十年代末至九十年代初，与当时欧美教育界对“标准化考试过度统治教学”的集中反思密切相关。以“为学习而评价”（assessment for learning）运动为代表，主张评价不应只是学期末的一次性统考，而应渗透在教学过程中，由学生与教师共同收集能够展示学习轨迹的材料——草稿、修改记录、自我反思、项目日志——并将其呈交为评价的对象。其核心理想是：让“学习过程”变得可见、可评价，从而解放那些被标准化测试所忽略的学习品质。

评价量表（rubric-based scoring）几乎是同时代的产物。它的基本逻辑是：传统评分之所以对“过程”不敏感，是因为评分标准太粗糙，只区分“对”与“错”。如果我们将评分标准拆得更细——比如给“论证的逻辑性”“解决方案的原创性”“对多种解法的尝试”分别设定维度与等级——那么统一考试也可以在保持信度的前提下，对更丰富的学习品质做出判断。

两段改良共享一个信念：评价看不见某些有价值的东西，是因为尺子上的刻度不够密。只要刻度足够密，看不见的东西就会变得可见。

为什么它们都撞上了同一面墙

本文并不主张这两类改良毫无价值。它们确实在特定限度内拓展了评价的敏感度。一个精心设计的几何证明评价量表，确实可以区分出“学生尝试了多种构造辅助线的路径”——这让那些多路径试探的努力至少不再与瞎做混为一谈。档案袋确实让教师有机会看到学生在一段时间内写作能力的成长弧线，而非只看到最后一次考试的那篇作文。这些都是真实的进步。

但本文必须指出它们的共同边界：它们都仍然在“评价”的制度框架内运作，而只要还在这个框架内，它们就无法逃脱那个构造前提——评价所认可的证据，必须是已经完成的、可以被第三方复核的、固定在某种可存档载体上的认知产物。

这个前提对档案袋评价造成的扭曲，比它表面看起来要隐蔽得多。档案袋要求学生在提交时，不仅提交最终成果，也提交过程性材料——草稿、修改记录、反思日志。在完全没有分配后果的情况下，这些材料可以保持一定程度的“粗糙”和“真实”：一个学生可能在反思日志里写“这个论证顺序我改了五次还是不满意，我甚至不确定自己到底有没有想清楚”——这段话没有经过润色，不具备展示价值，甚至读起来有点泄气。但现实是，档案袋在绝大多数实施情境中都被赋予了制度分量——即使开始的初衷并非如此。学校需要在制度上为档案袋寻找一个存身之处，而这个位置最容易找到的地方，就是分班参考、评优依据、综合素质评价的佐证材料。一旦档案袋与这些分配后果直接或间接地挂钩——哪怕只是“会放入学生个人档案，作为高校自主招生的参考材料之一”——过程性材料就立刻变了性质。它们从“给自己的记录”变成了“给评定者的证据”。这不是任何人的道德失误。这是制度逻辑：当我知道我的材料将来有可能被一个我不知道是谁的评定者翻看并据此对我做判断时，我对这些材料进行印象管理就不仅不是虚伪，而且是理性的。

学生很快学会如何管理这些证据——“草稿”要展示“修改轨迹”，什么样的修改轨迹看起来最认真？“反思”要展示“深度”，什么样的措辞读起来最有深度？这不是学生狡猾。这是制度教给他们的：你所提交的一切，都有被用来评判你的可能，你要为自己提交的东西负责。

而“生成性内歧义”恰恰是那种无法被预先管理的摸索——它不是学生在写的时候就知道这东西最终会成为一份好证据。在它刚发生的一刻，它看起来是混乱的、失败的、不确定的。如果学生知道这些过程性材料最终会被审阅，他就有充分的动力去整理那份混乱，删除那些不成功的试探，润色那些说不清的地方，把过程修剪成一段看起来专业的学习弧线。经过这一轮管理，档案袋里的过程性材料，已经不是生成性内歧义曾经发生过的那片现场了——它是那片现场被事后改造成的一座展馆。烟已经被冷却成固体，依照展柜的尺寸被切割好，贴上解说标签。档案袋的初始意图——让“学习过程”被看见——在这一刻转而产生了一个反讽效果：它让学习过程被看见，但被看见的是已经被策略性地管理过的过程。

评价量表的改良撞上的是同一面墙，只是路径不同。量表要在不同评分者之间取得可接受的一致性，就必须把评分标准翻译成可以被不同人独立执行的描述语。“尝试了多种解法”听起来不错，但如果不同评分者对于“什么叫尝试多种解法”的判断差异太大，这个维度在技术上就是不可靠的。为了提升信度，量表的设计者必须进一步细化：至少要写几条？不同程度的尝试是不是至少要呈现两页草稿才算“多种”？在操作压力下，评分细则会不自觉地滑向一套可计数的行为指标。这些指标一旦固定下来，“尝试多种解法”就从一种认知品质退化成了一种可被观察到的行为格式——学生很快就学会在草稿纸上多画几条线，多列几个标题，多写几行“思路”——不是在试探，而是在为“多种解法”这个评分维度制造可被计数的证据。

同样的逻辑也适用于“深度反思”。如果评量表把深度划分为三档——“能指出自己哪里卡住了”（浅）、“能分析卡住的原因”（中）、“能总结从卡住中获得的启示”（深）——学生就学会了在反思日志里用三段论的格式去呈现自己的反思过程，而不一定是真的有反思发生在了他身上。呈现和发生的裂隙，在量表出现的那一刻就形成了。

这里需要指出一个容易被忽略的区别：量表和档案袋虽然都捕捉不到生成性内歧义，但它们失败的方式不同，这个不同本身携带重要的信息。档案袋的失败，是事后管理——它捕捉到了那团东西的残骸，但残骸已经被整理过了。评价量表的失败，是事先定标——它只能看见那些自己事先规定了指标的行为，永远看不见指标之外的动作。这两条失败路径从不同侧面共同指向同一个结论：评价这个物种的构造前提，决定了它只能吃一种饲料——“已完成的、可存档的、可被第三方复核的认知产物”。生成性内歧义不是这个物种的食物。把量表磨得再细，把档案袋包容得再宽，改不了它吃什么。我们往食槽里放再多种类的草，也喂不活一头要吃肉的动物。

尺子构造前提的不可逾越性

至此，我们可以给出一个更为一般的论断。评价作为一种社会实践，其构造前提有三：第一，它必须产生可比较的判断（无法比较分高低的评价，在教育中就是无用的评价）；第二，可比较的判断必须依据可被第三方复核的证据，证据必须固定在某种可存档的载体上；第三，可存档的载体上，只能承载已经完成的、成形的认知产物。这三条前提不是任何一组改革者可以改写的“方案设计缺陷”——它们就是评价这个制度物种的基因。

而生成性内歧义的存在方式，与第三条前提正面相撞。它是一种正在发生、正在流动、正在成形过程中尚未凝定的认知动作。它的价值，不在于它最终产出了什么可以被打分的“正确结果”，而在于它在那段不确定的摸索中，锻炼了学习者那种“在没有人告诉你往哪走的时候自己摸路”的认知能力。这个锻炼的过程，是无法被物证化的——你可以事后追认它，复述它，但你无法在它还在发生的时候把它直接呈交为一组可被评分的证据。

这正是为什么，任何“更精细的评价维度”都无法从根本上解决这个问题。它不是刻度不够密的问题——它是尺子这种工具本身的构造前提决定了，它只能量已经凝成固体的东西。而生成性内歧义，是还在冒烟的流动。尺子再精确，也量不了烟。那些试图在评价框架内部捕捉它的改良，最终做的事情，要么是把烟冷却成固体再量（档案袋的事后整理），要么是根据烟冒出来的形态去推断它的成分（量表推断）——无论哪种，都不可避免地失去了烟之为烟的那个核心特性。

承认这一点既令人沮丧，也令人清醒。它意味着，并非每一件在教育上有价值的事，都能在评价制度的框架内找到一个安放自己的位置。有些事之所以有价值，恰恰是因为它们只能发生在那个框架的外面。而教育者不应因为“无法评价它”就否认它的存在权利。接下来的问题就变成了：在评价制度的外面，是否有可能存在一种实践，它不是试图去测量生成性内歧义——它只是为它的发生留出空间？

五、不被交出的叙事：在评价的外面，能有一块自留地吗

上一节的结论承认了评价制度的基因边界。这一承认如果诚实接受，紧接着的问题必须是是可以操作的：在评价制度的外面，具体可以做什么？这不是一句“教师要有爱心”或“要尊重学生的主体性”可以回应的问题——那些话是必要的背景态度，但不是具体的实践。在笔者所了解的一间初中语文课堂里，任课教师做过这样的尝试，它粗糙、朴素、没有名字，但它包含了本文所追问的那种实践的全部要件。

一间教室里的尝试

这位教师每学期会拿出两节课——期末前一周的某个下午，不是班会课，不是写作课，课表上写的是“阅读交流”。她发给每个学生一张A4白纸，不划线，不打格。她说的原话大致是：“这学期你在哪件事上，自己和自己较过劲？想写就写下来。不交也行。写完不想留着，撕了也行。我不看，也不打分。你如果想让我

看，可以放在讲台上那个盒子里，我会看，但我不写评语。你如果不想让我看，写完自己收着，期末结束之后你自己决定怎么处理。”

这个提议的每一条设定，在做教育研究的人看来都极不专业。没有评价量表，没有打分维度，没有前后测，没有档案留存，没有优秀案例展示，没有期末总结里“本学期我学会了什么”的固定句式。但恰恰因为它放弃了这一切制度性的装备，它创造了一种在本校其他课堂上几乎无法创造的极低压空间。学生写的东西不是用来展示学习成果的——他们自己决定什么值得写。有学生写了期中考试那道他卡了半个小时、最后画了一条完全不在原图里的线然后解出来的几何证明，“我当时差点哭了，我觉得我自己真的变聪明了一点”。有学生写了“整个学期我都在装，装我听懂了，装我不在乎，我其实很怕被老师提问”。有学生画了一整页乱七八糟的箭头和框线，旁边用小字标注“这是我的脑子在想一道题的路线图”。也有学生一个字都没写，把白纸折成飞机，下课偷偷飞进垃圾桶里——没有人追究他为什么不写。第二学期、第三学期，这个尝试继续保留了下来。不是作为一项“被推广的教改成果”，而是作为这位教师自己心里认定的一件“做了不会错”的事，安静地嵌在她的教学节奏里。

这个微型实践为本文提供了一个极其重要的参照点。它验证了：在评价制度的外面，真的可以有一块不自欺的缝隙。但要将它从一个感人故事提升到严格论证的位置，必须正面回答一个最难的问题：写白纸这件事，难道不是另一种形式的表演吗？学生难道不会揣摩“老师想看什么”，然后交上一篇漂亮的、自我解剖式的成长叙事？

表演的不同压力等级：为什么极低压可以做不一样的事

任何在人际空间中发生的表达行为，都无法完全隔绝表演性。Goffman（1959）在《日常生活的自我呈现》中已经论证过：一个人只要知道自己的言行可能被他人观看，就会对观看者预期做出某种程度的调适。如果“彻底消除表演”是这种实践成立的前提，那它从一开始就不成立。但表演有不同的压力等级。在升学档案里写一篇“综合素质自我评价”，它的读者是招生老师，它的后果是录取与否——这是一个极高压力等级的表演情境，学生几乎没有不表演的空间。而在那张期末白纸上，学生面对的是一个不会被打分、不会被写评语、甚至不会被任何他认识的人看到的纸面。他写什么、怎么写、写不写，都没有任何制度后果。这个情境的压力被制度性地压到了最低。

更重要的是，“不展示”原则堵住了一个致命的裂口——范本的出现。在任何课堂展示活动中，一旦某个学生的反思被拿出来作为“写得好的例子”被分享，一个隐性的标准就立刻形成了：此后每一个学生下笔时，心里都会不自觉地拿那个范本的“深度叙事”来掂量自己写的东西够不够“好”。范本是自发的最大敌人，因为它把一种本应只对自己负责的动作，重新纳入了被比较的视野。那位教师说“我不看，也不打分”的时候，她不只是降低了后果的严重性，她还在制度层面做了两件事：撤销外部判断的通道，撤销可以被模仿的范本。

这并不意味着学生从此不再表演。某些学生可能仍然会在白纸上写一些他们觉得“老师看了会喜欢”的话——即使本打算不给老师看，写的时候心里还是会浮现一个想象的读者。但关键在于：在这个低压情境里，他不一定每次都表演，他也可以有一页突然写出一句不事雕琢的、不必向任何人证明什么的真话。那一页可能写得不好看，可能只写了三行，可能和上一页的叙事风格完全不同。它不恒定。但正是因为它不需要恒定，它才不必表演。而在更长的时间尺度上，当他在连续两三个学期里反复面对这张白纸，他知道这张纸不会害他——不会被比较、不会被归档、不会在某次家长会上被悄悄拿出来给父母看——他就有可能在某一页上突然不再表演了。这不是人性的升华，这是低压环境给真话让出的一个窗口。

不是回收，是第一次生成

需要警惕一种误解：把这个过程浪漫化为“让学生做回真实的自己”。本文拒绝这样一种本质主义的叙事：仿佛在评价制度之下，埋藏着一个完整的、未被玷污的本真自我，只等着一次善意的邀请就可以破土而出。自我从来不是这样运作的。自我是在一个人与他所处环境中的各种制度、实践、关系的持续互动中，一点点被生成、被修改、被重新定向的东西。一个在全日制评价环境中度过了多年学龄的学生，他的认知习性、他的注意力方向、他对“什么值得做”的直觉，都已经被这套环境深刻地地形塑过。他并没有一块绝缘于制度的飞地可供返回。极低压空间能做的，不是在制度的铁幕上凿开一个洞，让他爬出去做回“本来的自己”——它做的是另一件事：在制度的日常运转的缝隙里，刻意设置一种情境，让那些在高压下自动被筛掉的、柔软的、不确定的、说不周全的认知经验，获得一个被自己重新看见的机会。这个机会本身，就有可能松动他被制度长期压实的某些注意力倾向，让他的自我在某个未被预设的方向上，长出一小截——或许只是一小截——之前不曾有过的分支。

这不是“回收主权”。“回收”预设了主权原本就属于他、只是被夺走了。而事实上，“对自己认知经验的意义进行叙述”这项权利，对他来说从来就不是一个天然就拥有过的东西。在制度化学习的漫长历史中，这项工作几乎全由制度代劳了——成绩单、教师评语、学期总结、升学档案，这些制度文本早已为他准备好了一套完整而连贯的关于“你学到了什么、你是怎么学的、你是什么水平”的叙事。他不需要自己来叙述，他只需要认领制度替他写好的那个版本。而白纸这件事——当他开始自己写，而不是等着成绩单来替他总结——他不是“回收”什么曾经拥有过的东西。他是在生成一种新的自我关系：一种“我自己来叙述我经历了什么，我的叙述不需要匹配制度档案里的那个版本”的关系。这种关系不是天然就有的。它是在这个具体的、低压的、非分配的书写行为中被第一次发生出来的。这就是为什么，本文坚持使用“生成”而非“回归”来描述这个过程的性质。

这个实践的防护条件及其脆弱性

这样一种实践，不可能也不应该被设计成一套“可推广的制度方案”。它是微型的、局部的、依赖具体的师生信任关系的。它不需要被写入学校发展规划，不需要被纳入督导评估指标。它只需要被保护——不被收编、不被展示、不被考核。

具体而言，它至少需要三样刻意的防护。其一，不设截止日：学生可以在学期任何时候写下，从来不写也没关系——因为一旦设了截止日，“在期限前必须完成”的压力就会渗透进来，有些学生会为了“交了差”而写一些他自己都不觉得重要的东西。其二，不许被展示：不在家长会、开放日、教研活动中被拿出来作为“我们班做得好的案例”示范。因为一旦被示范，“好的白纸长什么样”的范本就会出现，低压就被击穿——此后每一个学生下笔时，心里都会不自觉地拿那个范本来掂量自己。其三，不做前后比较：不能因为学生上一回写得“有深度”，这一回写得平淡，就暗示他“退步了”。他的每一次书写都是独立的。这三种防护的实际目的，是让这件事在制度层面变得“无用”——不产出任何可以用于外部比较、展示、考核的材料。而恰恰因为它在制度层面无用，它在学习者的自我关系层面，才可能重新变得有用。

一个最冷的质疑：如果连这个也被表演了呢

必须正面处理一个不可回避的质疑。有人会说：你费尽心思保护这块自留地，但学生难道不会在写白纸的时候，在心里暗暗揣摩“老师喜欢看什么样的反思”，然后交上一篇漂亮的、自我解剖式的、充满成长弧光的叙事？而且更极端地——如果他连续三年都在写白纸，他会不会逐渐习得一套“写白纸的格式”，让自己在白纸上也越写越像一份档案袋材料？

这是一个诚实到无法被技巧绕开的反驳。本文的回应分两步。

第一步，承认这个风险是真实的。任何低压空间，只要持续存在足够长的时间，都有可能被参与者逐渐适应并开始在其中进行策略性管理。白纸也不例外。它不是一个魔法结界，一旦设立就从此免于被制度逻辑渗透。它只是一块在制度的铁幕上刻意留出的低压区——这块低压区的维持，需要教师持续地不去利用它、不去展示它、不去考核它。这本身就是一种持续的制度性克制。一旦这份克制松懈了——比如教师某次忍不住在家长会上提到“我们班有些孩子写得特别感人”——白纸的极低压条件就被破坏了，它被拉回了可比较的引力场。

第二步，但即使白纸被部分地表演化了，它依然和档案袋中的“反思日志”有一个根本的不同：白纸不承担任何制度分量。它不是综合测评的佐证材料，不是评优加分的数据来源，在任何外部评定者面前它根本不存在。一个在白纸上写出“看起来很美”但其实只是在模仿“好反思格式”的学生，他并没有因此从制度那里获得任何奖励。他写在白纸上的东西，与写在作文本上的东西，在他自己的经验结构里占据完全不同位置——作文本上写得好会被表扬；白纸上写得好，没有后果。没有后果，就意味着表演的动机被大大削弱了。因为表演总是为了某个观众。当观众被制度性地撤销，表演就只剩下一个给自己看的“自我表演”——而那恰恰是另一个名字：尝试弄清楚自己想说什么。

六、结论：在可被编目的学习之外

本文的论证从“不敢犯错”这个流行的诊断出发，最终停在了一间教室里的一个朴素动作上。它不是要提供一个解决评价难题的制度方案——这类方案已经太多，而它们共同的遗憾是，每一套方案都试图在“评价”这个大标题下做文章，结果无一例外地被那个标题的引力场捕获。本文想做的，是把这个引力场本身的边界指认清楚，然后指认边界之外的一小片空地——它小，它不产量化产出，它无法被任何督导指标覆盖，但它在那里。

本文的核心主张可以凝为以下四点。

第一，在标准化评价长期塑造的学习环境中，学习者所经历的深层变形不只是“不敢犯错”，而是一种更早的自律——“生成的非法化”。他不再需要制度告诉他什么不准想——他的注意力已经被训练得不再往那些“不成形的东西”上流了。这套自律的发生，不是经由三个独立原因的先后作用，而是在日常教学事件的交织运作中——采分点窄化作为背景辐射、时间与反馈技术架空“扛住不确定”的意愿、以及延迟扣分将不确定感条件化为负价信号——三者在同一个作业周期中彼此渗透、互相喂养，像绞紧的绳股一样一道一道地被夯进学习者的认知身体。

第二，被这套自律筛掉的东西中，有一类具有独特认知价值的——“生成性内歧义”。它不是能力不足，不是胡思乱想，而是一个人在现有概念网络边界处，用不成熟的试探去触碰新结构时必然经历的、无法归为对或错的认知状态。这种状态与“生产性失败”文献（Kapur）所描述的探索活动在认知品质上有深层交集——都是在缺乏先备知识的情况下进行生成性试探——但它与Kapur实验情境的制度条件有本质不同：它不是在“免于评价后果”的庇护所里发生的被设计好的学习策略，而是在真实课堂的日常评价制度中不断被压缩、因而趋向于凋亡的认知本能。它与技术性失误在标准化评价的视力范围内共享同一个“未得分”的外观，而这一不可区分性是结构性的——它源于评价制度必须依据已完成的、可物证的认知产物来进行判断这一构造前提。认知失衡（Piaget）提供了失衡是认知发展驱动力的机制，productive failure（Kapur）提供了探索比直接讲授更促进深层学习的实证判断，本文在此地基上接过了这两个判断，然后追问了一个它们没有追问的问题：在这种失衡和探索被持续地、无声地标记为“不划算”的制度条件下，学习者会学到什么？

第三，任何在“评价”框架内试图通过细化指标来保护这一状态的改良，都受限於上述构造前提。档案袋评价与评价量表在特定限度内有效，但它们的共同边界在于：一旦过程性材料被赋予制度分量，它们就会被学习者策略性地管理，从而失去“正在发生的认知动作”的本貌。尺子再软，也量不了还没凝成形的烟。

第四，本文所捍卫的实践，不是一套新的评价维度，而是刻意留出的一种“不被交出的叙事”。它的价值不在被制度认可，而在于学习者在制度化的学习生涯中，得以在某个不被考核、不被展示、不被比较的缝隙中，重新生成一种不服务于外部标准的自我关系。这种关系不是对“本真自我”的回归——不存在一个等待被解放的先在本质——而是在评价制度的缝隙里，让一种不同于“被成绩单和评语栏编目的自我”的主体性形态，得以被第一次发生出来的条件。

最后，本文以如下方式给出其核心判断的证伪条件。

第一，关于“生成的非法化”的发生学判断——即自我筛查倾向是通过采分点窄化、不确定性不核算化、延迟扣分的条件化在三者的交织运作中被发生出来的——如果在针对不同类型学校的课堂观察或学生访谈中出现系统性的反例：在高度应试压力环境中，有稳定且大量比例的学生并未表现出注意力从模糊念头自动撤离的倾向；或在几乎没有应试压力的环境中，学生仍然大面积表现出同等的自我筛查，那么，本文所提出的机制链条就应当被修正——它可能夸大了评价制度相对于其他因素（如认知发展阶段本身的限制、数字媒介注意力模式等）的解释力。

第二，关于极低压书写能在足够长的时间尺度上生成不服务于外部标准的自我关系这一判断——如果在那些严格维持无截止日、不展示、不做前后比较的课堂实践中，出现压倒性的长期追踪观察证据表明：绝大多数学生在连续多个学期后，其书写内容在完全撤除教师观看权（学生写完即自行处置，不交任何人）的条件下，仍高度趋同于制度性反思的套路格式，并且这一模式在增加时间跨度后仍未发生任何实质性变化——那么，本文关于“在低压情境下可以生成不被制度编目的自我认领”的判断，就应当被修正。在那样的证据面前，我们或许不得不接受一个更冷的结论：与现代学校制度耦合至深的评价文化，已经彻底压倒了学习者在任何情境下生成不服务于外部标准的自我叙事的可能性。

但在那样的证据出现之前，我们仍有理由为一间间具体的教室、一个个具体的教师，争取那块可以被保护住的、不大却真实的教育缝隙。因为它所庇护的，是一个人在漫长的制度化学习生涯中，为数不多的、可以不为了被任何人看见、而只是为了让自己成为自己历史的叙述者，而拿起笔的那个时刻。

七、最近邻辨析：生成性内歧义不是阈限态

本文的核心区分——生成性内歧义与技术性失误在评价的视力范围内共享同一个“未得分”外观——必须正面回应一个最贴的占位：迈耶与兰德（Meyer & Land, 2003, 2006）的阈限概念与“麻烦的知识”（另见 Perkins, 1999）。阈限概念理论指出，学习者在跨越一个关键概念门槛时，会长时间卡在一种阈限态（liminal space）里：似懂非懂、反复卡壳、外观上与“没学会”难以分辨。这与生成性内歧义几乎一眼相似。

本文承认这一相似，但剩余分工在于两者所描述的对象根本不同。阈限概念理论描述的是学习者的认知发展阶段——它是一个学习者在掌握某个门槛概念的过程中必然经历、并最终将穿越的时间性相位，其分析落点是学习者内部的认知状态。而生成性内歧义的主张不是“学习者会经历一个困惑期”，它是关于评价制度的一个结构性判断：评价制度作为一个必须依据可物证的已完成产物做判断的物种，在其视力范围内，结构性地无法把生成性内歧义与技术性失误分开——因为二者都只呈现为“未得分”这一个证据形态。阈限理论谈的是学习者穿越门槛的发展轨迹，生成性内歧义谈的是评价装置的结构性感官。前者可以靠更好的教学支持来陪伴穿越，后者无法靠更精的评价技术来分辨。

由此得到可裁决差别：若存在一种足够精细的过程性评价，能在不改变评价之为评价（依据可物证产物做可比较判断）的前提下，可靠地把生成性内歧义与技术性失误分开，则本文的"结构性不可区分"主张失效，它退化为一个评价精度问题。本文预测不存在这样的评价——因为任何评价要做出判断，都必须把"仍在发生中的认知动作"截停为一个可物证的产物，而这一截停恰恰抹去了区分二者所需的、唯一存在于过程之中的信息。

参考文献

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Broadfoot, P. (1996). *Education, assessment and society: A sociological analysis*. Buckingham: Open University Press.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Random House.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer Press.

Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.

Jackson, P. W. (1968). *Life in classrooms*. New York: Holt, Rinehart and Winston.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45–83.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York: International Universities Press. (Original work published 1936)

Seligman, M. E. P. (1975). *Helplessness: On depression, development, and death*. San Francisco: W. H. Freeman.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281–294.

Meyer, J. H. F., & Land, R. (2006). *Overcoming Barriers to Student Understanding: Threshold Concepts and Troublesome Knowledge*. London: Routledge.

Perkins, D. (1999). The many faces of constructivism. *Educational Leadership*, 57(3), 6-11.

修订说明

本文经德麦国际 SDE 学派编辑（Claude）依两轮修订体例打磨：补第七节，与阈限概念／麻烦的知识（Meyer & Land, 2006；Perkins, 1999）正面切割——阈限态谈的是学习者的认知发展阶段，生成性内歧义谈的是评价装置的结构性色盲，二者对象不同；给出'是否存在能可靠分辨二者的过程性评价'的可裁决差别。作者原有生成性内歧义、自我筛查语法与微型实践定位未改。作者的核心判断、论证结构与原初命名均未改动。