

还在动的时候就读了数

论课堂判定的时点相对于重组时长过早，如何把仍在进行的变化冻在当时的构型上

物理学 × 心理学 × 工程学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅 约1.9万字 · 27页

发表 2026年8月3日

摘 要

课堂每隔一两周作一次「会了没有」的判定，并据此配置此后的一切条件；判定的时点由课表决定，不由被判定的那件事本身决定。物理学主张决定一个系统有没有在变的是弛豫时长与观测窗口之比，心理学主张一次提取会主动抑制同类竞争项，工程学主张进不了传感回路的状态就进不了控制。三家争的是变化由什么驱动，而共同假定了一条没说出口的话：**正在变的部分与不在变的部分，可以在观测的那一刻被分开**。推翻它的材料来自物理学自己——**物理老化**：被判为「冻住了」的玻璃态系统，性质随冻住之后所等的时间持续演化，且演化的起点正是被冻住那一刻的构型。⇒ **读数为零是关于窗口的，不是关于系统的**；而判定这个动作参与了后续演化的初始条件。由此命名**冻结构型**，并给出一个不需要分母、任何学校都做得起的读数：**双向翻转率**——同一份卷子隔四周原样再做、期间不教，看有多少人翻过来、两个方向各占多少。**单向翻转是遗忘的签名**；「不会到会」那一支显著**非零**，是本文的独有签名，遗忘与「没学扎实」都预测不出它。文中另逼出教育测量学的一处自相矛盾，并给出一段初中数学单元的完整推演：同一次不及格里混着重组时长差一个数量级的三类学生，而补救内容的性质决定了哪一类被救、哪一类被冻。

这一篇由哪三个学科的理论体系撞成

第一家主张一个系统是固体还是流体、有没有在变，只由弛豫时长与观测时长之比决定，与驱动力大小无关；第二家主张决定什么留下的是提取时的竞争关系，遗忘不是衰退而是提取这个动作自己造成的抑制；第三家主张系统实际表现成什么样，取决于哪些状态进得了反馈回路，而非这些状态本身多要紧。三条不能同时为真：**工程学「不可观测就证明它稳定然后不管」的日课，正是物理学诊断的病根——看起来不动恰恰不是稳定的证据**（物理判工程）；心理学最有力的教学建议是用提取来巩固，而它同一套实验证明提取会主动抑制同类未被提取项（心理判自己）；而物理学的处方是把窗口拉长到超过弛豫时长，这在课时有限的课堂里不可执行，且按工程学，推迟判定等于降低采样率，只会加重混叠（工程判物理）。三家均为站外的公开文献，链接直达原始出处，可自行核对。

物理学 · 玻璃转变、德博拉数与物理老化

[Debenedetti & Stillinger, Supercooled Liquids and the Glass Transition \(<i>Nature</i> 410, 2001\)](#); 另见 [Reiner, The Deborah Number \(<i>Physics Today</i> 17:1, 1964\)](#); [Struik, <i>Physical Aging in Amorphous Polymers and Other Materials</i> \(Elsevier 1978\)](#)

主张玻璃转变不是相变，而是弛豫时长超过了实验时长、系统被困在能量地形的一个盆地里；它并没有停止，只是慢到读不出来。这一支自曝：被判为冻结的材料，其性质随**等待时间**持续演化，蠕变曲线形状不变只沿时间轴平移——它用这条预测保质期，从没有人拿它当作「正在变与不在变可否当下区分」的证据。

<https://doi.org/10.1038/35065704>

心理学 · 提取诱发遗忘

[Anderson, Bjork & Bjork, Remembering Can Cause Forgetting \(*J. Exp. Psychol.: LMC* 20:5, 1994\); 另见 Anderson & McCulloch, Integration as a General Boundary Condition \(*JEP:LMC* 25:3, 1999\)](#)

主张练习一部分内容会**主动削弱**同类中未被练到的那些，遗忘是提取这个动作自己造成的抑制。这一支自曝一处边界条件：当类别内成员被整合成相互关联的结构时，该效应消失甚至反转——它用这条划定适用范围，没有用来问一次判定本身在改变什么。

<https://doi.org/10.1037/0278-7393.20.5.1063>

工程学 · 可观测性、采样定理与抗混叠

[Kalman, Mathematical Description of Linear Dynamical Systems \(*SIAM J. Control* 1:2, 1963\); 另见 Shannon, Communication in the Presence of Noise \(*Proc. IRE* 37:1, 1949\)](#)

主张不可观测的状态无论多要紧都进不了反馈回路，工程实践对这一块只做一件事：证明它稳定，然后不管。这一支自曝采样定理的另一半：采样过疏得到的不是「没有信号」而是**混叠**——一个虚假的、比真实变化慢得多的低频信号；对策不是提高采样率，是先滤掉快成分、宁可承认「这部分我不测」。

<https://doi.org/10.1137/0301010>

目 录

摘 要	5
Read While It Was Still Moving.....	5
一、问题从哪里来	6
二、物理学的单因主张：已有的样子与观测条件不容，驱动我们对路径的判定改变	6
三、心理学的单因主张：已有表征与提取路径不容，驱动可及条件改变	7
四、工程学的单因主张：控制路径与传感条件不容，驱动系统实际表现改变	8

五、三家为何不能同时为真	8
六、共有前提，与推翻它的那件材料.....	9
七、冻结构型	9
八、冻结构型的四种亚型	10
九、二维辨别格	11
十、读数与一句不含判断词的问法	11
十一、十处兑现	12
十二、一个班走完：从判定到配置到锁死.....	13
十三、三家为何各自到不了	13
十四、为什么常见的改革处理不了这一件.....	14
十五、与最接近的几种说法的分界	15
十六、一处被逼出来的自相矛盾	16
十七、与同批另两篇的分工	17
十八、五类场景中的窗比，与其中两处收窄.....	17
十九、三个领域反过来加上的限制	18
二十、不适用的边界	18
二十一、经验研究设计	19
二十二、证伪条件与轮次预测	20
二十三、判定可以怎么改：三个动作.....	21
二十四、反噬预言与伦理限制	21
二十五、自反	22
二十六、分层引用，与三个本文解释不了的洞.....	22
二十七、三条本文预测而尚无人测的现象.....	23
二十八、读数的可比性：三处本文自己交代不清的地方	24
二十九、一处必须说清的伦理前提：谁有权要求复测.....	24
三十、结论与一条写死日期的赌注	25
注 释.....	25

论课堂判定的时点相对于重组时长过早，如何把仍在进行的变化冻在当时的构型上

物理学、心理学与工程学的三重动力解释

摘要

课堂每隔一到两周就要作一次"会了没有"的判定，并据此配置此后的一切条件——补救、分层、跳过、放弃。判定的时点由课表决定，不由被判定的那件事本身决定。物理学主张决定一个系统"有没有在变"的是弛豫时长与观测窗口之比，心理学主张一次提取会主动抑制同类竞争项，工程学主张进不了传感回路的状态就进不了控制。三家争的是变化由什么驱动，而共同假定了一条没有说出口的话：**一个系统里正在变的部分与不在变的部分，可以在观测的那一刻被分开**。推翻它的材料来自物理学自己——物理老化：一个被判定为"不再变化"的玻璃态系统，其性质随被冻住之后所等的时间持续演化，且演化的起点正是被冻住那一刻的构型。带回课堂便得到本文命名的对象：**冻结构型**——一次判定在被判定者身上留下的、并非关于内容而是关于"当时重组进行到哪一步"的状态；此后的慢速变化都从这里出发，而它不进入任何读数。承重判断是：学习之所以少，不是它没有在进行，而是课堂的观测窗口短于重组时长，因此系统性地把慢于窗口的变化读作零，并据此把它冻在当时的构型上。本文给出窗比与双向翻转率两个读数、一张二维辨别格、一句不含判断词的问法、与九种相近说法的分界、一处强制性的自相矛盾、三处反过来加上的限制、七条证伪条件与一条写死日期的赌注。

关键词：冻结构型；窗比；双向翻转率；物理老化；判定时点

Read While It Was Still Moving

On classroom judgements taken earlier than the reorganisation they measure, and the configurations they freeze

Abstract. Every week or two a classroom decides whether each learner "has got it," and configures everything downstream on that decision. The timing of the decision comes from the timetable, not from the thing being decided. Physics holds that whether a system is changing is settled by the ratio of relaxation time to observation window; psychology holds that retrieval actively suppresses competing items; engineering holds that a state which cannot reach the sensing loop cannot reach control. The three disagree about what drives change, yet share an unstated premise: that within a system, what is changing and what is not can be separated at the moment of observation. The material overturning this comes from physics itself. Physical ageing shows that a glassy system pronounced unchanging continues to evolve with waiting time, from the configuration it held when it was frozen. Carried back to

the classroom, this yields the object named here: the **frozen configuration** — what a judgement leaves in the person judged, not about content but about how far a reorganisation had got; all subsequent slow change departs from it, and it enters no measure. The load-bearing claim is that learning is scarce not because it is absent but because the observation window is shorter than the reorganisation, so slow change is systematically read as zero and frozen at that reading. The paper supplies two measures — window ratio and two-way reversal rate — a matrix, a modality-free question, separations from nine adjacent accounts, one forced inconsistency, three constraints imposed in return, seven falsification conditions and a dated wager.

Keywords: frozen configuration; window ratio; two-way reversal; physical ageing; timing of judgement

一、问题从哪里来

一个学生在单元测里做错了三道同类题，被判为“没掌握”，进了补救组。补救组的任务更简单、步子更小、示范更多。四周之后，如果有人拿同一份卷子让他再做一次——而这四周里没有人再教过这部分——他有相当的可能做对。没有人会去做这件事，因为判定已经作出，条件已经按判定配置，那三道题早已从他的课程里消失了。

反过来的情形同样常见：当堂全对、被判为“已掌握”的学生，四周后同一份卷子做不出来，而这四周里同样没有人再教过。

两种情形通常被分别归因：前者叫“开窍了”，后者叫“忘了”。两个词都指向学生，都不指向那次判定。而它们放在一起时露出一件更麻烦的事：**判定翻转是双向的**。若只是遗忘，翻转应当只有一个方向；能从“不会”翻到“会”，说明在判定与复测之间，没有教学的情况下，有东西还在进行。

本文追问的正是这一段：课堂的观测窗口有多长，被观测的那件事需要多长，两者的比是多少，以及当这个比小于一时，一次判定到底做了什么。

本文不报告新采集的数据，是一篇理论性文章；其经验主张需要预注册研究与独立复核检验。

二、物理学的单因主张：已有的样子与观测条件不容，驱动我们对路径的判定改变

凝聚态物理处理过一个形式上完全相同的问题。一块玻璃是固体还是液体？答案不由它的组成决定，也不由驱动力大小决定，而由**弛豫时长与观测时长之比**决定。这个比在流变学里有名字，叫德博拉数

(Reiner, 1964)：比值大，物质表现为固体；比值小，表现为流体。同一块沥青，在几秒的观测里是固体，在几十年的观测里会滴下来。

玻璃转变本身也不是一次相变，而是**弛豫时长超过了实验时长**，系统被困在能量地形的一个盆地里，无法在观测期内探遍它本该探遍的构型 (Debenedetti & Stillinger, 2001)。它并没有停止，只是慢到了读不出来。

这一支的单因主张是：**决定一个系统"有没有在变"的，是弛豫时长与观测窗口之比，不是驱动力的大小。**时序读数是：先有观测窗口的设定，随后同一系统被判为不同的物态，物态判定又决定用哪一套方程去描述它此后的路径。

单因锁定成立：这一支不认为"变化速率"与"观测窗口"是并列因素。它主张在给定观测窗口下，物态是被这个比唯一决定的。

它自己知道、却不往这个方向用的事实是**物理老化**：一块被判定为"冻住了"的玻璃，其密度、模量、蠕变顺应性**随冻住之后等待的时间**持续变化，而且变化的形态由被冻住那一刻的构型决定 (Struik, 1978)。这一支用它来预测材料的长期性能与保质期；它从来不被当作一条关于"正在变与不在变可否当下区分"的证据。

三、心理学的单因主张：已有表征与提取路径不容，驱动可及条件改变

记忆研究提供第二个位置，且它的主张与直觉相反。练习一部分内容，不只是让这部分变强，还会**主动削弱同类中未被练到的那些**。这个效应有稳定的实验范式与三十年的复制记录 (Anderson, Bjork & Bjork, 1994)：从一个类别中反复提取若干成员，会使同类中未被提取的成员在后测里显著低于从未接触过该类别的基线。

这一支的单因主张是：**决定什么留下的，是提取时的竞争关系；遗忘不是衰退，是提取这个动作自己造成的抑制。**时序读数是：先有一次提取，随后同类竞争项的可及性下降，可及性的改变又决定下一次提取会调出什么。

单因锁定成立：这一支明确主张抑制是提取的直接后果，不是注意力、时间或干扰的副产品，并用独立线索测试排除了这些替代解释。

它自己知道、却不往这个方向用的事实是这一效应的**边界条件**：当学习者把类别内的成员整合成一个相互关联的结构时，提取诱发遗忘会消失甚至反转——提取一个成员反而促进同类 (Anderson & McCulloch, 1999)。这一支用它来划定效应的适用范围；它没有被用来问：**当一次判定要求学习者提取某几项时，这次判定本身就在改变其余项的可及性——那么被判定为"不会"的那些，究竟是本来不会，还是被这次判定压下去的。**

四、工程学的单因主张：控制路径与传感条件不容，驱动系统实际表现改变

控制工程提供第三个位置。一个系统能否被控制，不只取决于有没有执行机构，还取决于它的内部状态能否从可测的输出被重建。不可观测的状态无论多重要，都进不了反馈回路（Kalman, 1963）。工程实践据此把系统分解成可控可观、可控不可观、不可控可观、不可控不可观四块，**并对最后一块只做一件事：证明它是稳定的，然后不管。**

这一支的单因主张是：**决定一个系统实际表现成什么样的，是哪些状态进得了回路，不是这些状态本身有多要紧。**时序读数是：先定传感与采样方案，随后可观测子空间被确定，回路只在这个子空间里工作，系统的实际行为随之被塑造成回路看得见的那个样子。

单因锁定成立：这一支不认为"重要性"能替代"可观测性"。一个对性能至关重要而不可观测的状态，在设计上与不存在等价。

它自己知道、却不往这个方向用的事实是采样定理的另一半：若采样频率低于信号最高频率的两倍，得到的不是"没有信号"，而是**混叠**——一个虚假的、比真实变化慢得多的低频信号（Shannon, 1949）。工程实践的对策是在采样之前加抗混叠滤波器，把快于半采样率的成分先滤掉，宁可承认"这部分我不测"，也不让它伪装成慢变化。这一支用它来保证信号重建的正确性；它没有被用来问：**一个采样过疏的评价体系读到的进步曲线，是不是一条混叠产物。**

五、三家为何不能同时为真

三家把决定性的驱动放在三个不同位置：物理学放在已有构型与观测条件的不容，心理学放在已有表征与提取路径的不容，工程学放在控制路径与传感条件的不容。并列起来会得到一句无风险的话：评价要及时、要准确、要覆盖重要内容。这句话正确，且任何失败都可以事后归入其中一条。

三家实际上互相取消。

工程学的日课正是物理学诊断的病根。工程的标准做法是：不可观测的那一块，证明稳定即可不管。而按物理学，"看起来不动"恰恰不是稳定的证据——它可能只是弛豫慢于窗口，而被冻住的系统正在老化。若物理学那一条是对的，工程学对不可观测子空间的整个处置方式在原理上不成立，因为它用来判定"稳定"的观测手段本身就是那个窗口。

心理学的日课正是它自己证明会造成损害的那件事。这一支最有力的教学建议是用提取来巩固学习；而它同一套实验证明，提取会主动抑制同类未被提取项。若把判定视为一次大规模提取，则每一次判定都在削弱它没有考到的那些内容。工程学会说这是采样的必要代价，物理学会说这说明观测本身改变了系统——三家在"能不能一边测一边不动它"这件事上不可调和。

物理学的处方在另两家眼里不可执行。物理学的办法是把观测窗口拉长到超过弛豫时长。课堂无法这样做：课时有限，判定必须在单元结束时作出，而按工程学，推迟判定等于降低采样率，只会加重混叠。

三家必须有两家错，或者三家的病灶都放错了位置。本文主张后者。

六、共有前提，与推翻它的那件材料

三家争的是**变化由什么驱动**。而它们共同假定了一条从未说出口的话：

> **一个系统里正在变的部分与不在变的部分，可以在观测的那一刻被分开。**

物理学问变化沿什么路径进行，心理学问什么被记住什么被抑制，工程学问哪些状态进得了回路——三种问法都要求先有一个"此刻在变／此刻不在变"的分类。把这句话念给三家听，三家都会说这还用说吗。

推翻它的材料在物理学自己手里，而且是材料工程的常识：**物理老化**。一块被判为已经冻结的玻璃态材料，在恒温恒压、无任何外加驱动的条件下，其性质随**等待时间**持续演化；蠕变曲线的形状不随等待时间改变，只沿时间轴平移，平移量是等待时间的幂函数（Struik, 1978）。这意味着三件事，逐件都与共有前提相反：

其一，**被判为"不在变"的系统正在变**。它变得慢，慢到在作出判定的那个窗口里读数为零。

其二，**它此后的变化以被判定那一刻的构型为起点**。冻住的时刻不同，往后的老化轨迹就不同——**判定这个动作参与了后续演化的初始条件**。

其三，**读数为零是关于窗口的，不是关于系统的**。换一个更长的窗口，同一个系统会给出非零读数；这不是测得更准了，是测的本来就是另一个量。

带回课堂，共有前提塌掉。一次判定不是对一个已完成状态的读取，它是在**一段仍在进行的重组中间，取了一个点，并把这个点当作那个人的状态记下去**。而按老化，被记下的那一刻不是中性的：往后的一切慢速演化都从它出发，而课堂据它配置的条件又限定了这条演化能走多远。

七、冻结构型

据此本文命名一个对象。

> **冻结构型**：一次判定在被判定者身上留下的状态，它不是关于内容掌握与否，而是关于当时那次重组进行到了哪一步；此后不受教学的慢速演化以它为起点，而它不进入任何读数。

四个必要条件缺一不可。第一，被判定的那件事的**重组时长长于观测窗口**——学习者当时确实还在进行中。第二，判定被**作出并记录**。第三，**后续条件依判定配置**——分层、补救、跳过、免修、放弃，任何一项都算。第四，**判定与后续条件之间没有再判定**，因此那个构型在此期间无人过问。

三条性质使它区别于邻近的坏情况。

它不是遗忘。遗忘是单向的；冻结构型预测**双向**翻转，且"不会翻成会"的那一支不能由额外练习解释，因为期间没有教学。

它不是测量误差。经典测量理论假定存在一个在测量期间稳定的真值，误差围绕它波动；冻结构型说的是真值本身在以慢于窗口的速度移动，因此两次测量之间的差不是噪声，是位移。

它不是能力差异。同一个学生不同内容上的窗比不同——有的内容重组以分钟计，有的以月计——因此同一个人可以在一处被冻住、在另一处不被冻住。

需要排除两类伪阳性。其一，**判定不配置后续条件**：一次不计分、不分层、不改变任何安排的诊断性测验，即使窗比很小也不产生冻结构型，它只是一次无效读数。其二，**重组已经完成**：若该内容的重组时长确实短于窗口，判定就是有效的，与本文无关。

也要排除一类伪阴性：**推迟判定不等于没有冻结**。把单元测从第二周挪到第六周，若该内容的重组以月计，窗比仍小于一，只是冻得晚一点。延迟本身不是解药。

八、冻结构型的四种亚型

按被冻住的是哪一类重组，可以分出四种。四种的形成机理相同，而复测时的翻转方向不同——这使得亚型分类本身给出一条可查的细化预测。

表征重组冻结。学习者正在把一个问题从表面特征重述为关系结构：从"这是一道行程题"重述为"这是两个量按同一倍数变化"。这个过程以周计。判定落在中途，他被冻在"部分重述"上——遇到新题时用一半旧描述、一半新描述，做对的那些靠旧描述里恰好够用的部分。复测时这一类**最容易出现"不会到会"**，因为重述会在无教学的情况下继续进行。

策略选择冻结。学习者已经掌握两套做法，正在从"只会一套"过渡到"按情形选一套"。判定落在中途，他被冻在"有两套而不知何时用哪套"。这一格的表现极不稳定：同一类题时是对是错，取决于哪一套先被想起。教师最常见的处置是要求他"统一用老师教的那一套"——而这恰好把选择这件事从他手里拿走，使冻结变成永久。

判据重组冻结。学习者正在把"老师说对就是对"换成"我能自己验证"。这个过程以月计，中间必然经过一段**权威已被削弱而验证能力尚未建立**的时期。判定落在这一段，他被冻在半信半疑上。这一类复测时**最容易出现"会到不会"**——因为原来靠记忆与信任支撑的正确，在权威被削弱之后失去了支撑，而新的支撑还没长出来。这一格最容易被误读为"退步"，实际上它是四类里最接近成功的一类。

整合冻结。学习者正在把若干孤立结论连成一片：知道它们有关，说不出关系。这个过程时长跨度最大，从数周到数月。判定落在中途，他被冻在"能分别复述、不能一并使用"上。这一类在单项测验里全部通过，只在需要跨项的综合题上失效，因此它被单项判定系统性地漏掉。

四种可以并存，且同一个学生不同内容上处于不同亚型。因此**不能把一个人整体判为"被冻住了"**，必须按内容与重组类型分别记。这一条同时限制了本文的读数：双向翻转率是一个混合量，若不按亚型分层，

第一类的"不会到会"与第三类的"会到不会"会互相抵消，总翻转率可能接近零而两个方向都不为零。**分方向报告因此不是精细化要求，是这个读数能不能用的前提。**

九、二维辨别格

横轴取窗比：从该内容首次出现到判定作出之间的时长，除以该内容重组所需的时长。**纵轴取判定是否配置后续条件。**两轴相互独立。

	判定不配置后续条件	判定配置后续条件
窗比大（窗口长于重组）	冗余型：读数有效而无人使用；浪费，但不产生损害	对时型：读数有效且被使用，判定与被判定之物同步
窗比小（窗口短于重组）	观望型：读数无效但无后果； 关键对照格	冻结型：读数在原理上无效，而此后的一切条件 按它配置。本文的靶格

右下角是靶格，三条性质使它难以被发现。

其一，短期指标表现为改善。窗口越短、判定越频繁，即时正确率越高、进度感越强、"及时反馈"这一条越好看。任何以当轮读数评价的体系都会奖励缩短窗口。

其二，失效不产生信号。被冻住的那批人的继续演化不产生任何记录——他们在补救组里的表现由补救任务决定，而补救任务是按冻结结构型配置的，因此永远与判定一致。**判定的自我实现在这里不是比喻，是记录结构的必然后果。**

其三，它自我加固。判定失效表现为"这批学生基础差、需要更多跟踪"，标准处置是提高判定频次；频次一提高，窗比进一步变小。

左下角的观望型是关键对照格：**同样短的窗口，只要判定不配置后续条件，本文预测长期表现不受损害。**这一格的存在使本文与"频繁测验有害"划清界线——问题不是测，是测完之后按它安排了什么。

十、读数与一句不含判断词的问法

窗比需要两个量。分子是从该内容首次出现到判定作出的时长，从课表与教案直接可读。分母是该内容的重组时长，必须实测：在不作判定、不配置条件的条件下，跟踪同一批学习者的表现随时间的变化，取达到平台期所需时长的下四分位作为下界。分母难测是本文最大的一处未闭合，本文不掩饰这一点。

双向翻转率不需要分母，因而是本文更可用的一个读数：把同一份判定在两个时点各做一次，中间不教学，统计判定结果发生翻转的比例，并**分开报告两个方向**。单向翻转（会到不会）是遗忘的签名；**双向翻转，尤其是"不会到会"那一支显著非零，是本文的独有签名。**这个数只需要一份现成的卷子和四周的间隔，任何学校都做得起。

判据写成一句不含判断词的问话：

> 把这门课上一次作出"会了没有"判定的那份卷子，隔四周原样再做一次、期间不教，有多少人的结果会翻过来？两个方向各占多少？

五个场景跑一遍，答案互不相同：

场景	判定时点	该内容重组时长的量级	窗比
中学数学单元测	教学结束后 1-2 周	分数与比例的概念重组以月计（文献常见口径）	远小于一
编程训练营周考	每周	递归与指针的重组以周计，个体差异极大	接近一，方差极大
医学住院医的当场评估	操作当场	单项技能以次计	大于一
语言能力分级	按课程周期评	官方给出的相邻级别间指导学时以数百小时计（欧洲语言共同参考框架及其配套指南）	远小于一
工业技能认证	一次通过、长期有效	单项操作以天计	远大于一

第四格与第三格是查出来的，不是从命题推出的。语言能力框架的配套材料公开给出了相邻级别之间的指导学时，而多数语言课程按学期评级——**这一格的窗比可以直接算出来，而且算出来的数字与课程设计之间的落差是公开可核的**。住院医评估工具的时点设计同样是公开文件：当场评估与半年一次的里程碑评估在同一套体系里并存，两者的窗比相差两个数量级，而它们被同等地用于配置后续轮转。

十一、十处兑现

其一，德博拉数对应窗比，因此本文的核心读数与流变学的核心读数量纲相同，都是无量纲比，可以跨领域比较。其二，玻璃转变不是相变而是窗口效应，对应"没掌握"不是一种状态而是一个读数，因此研究必须报告观测窗口而不只报告结果。其三，物理老化对应判定后的慢速演化，因此本文预测复测差异随两次测量的间隔呈幂律而非线性，这一条可直接拟合。其四，老化曲线只沿时间轴平移而形状不变，对应被冻住的学习者其后续学习曲线形状与未被冻住者相同而起点不同，可用同一模型拟合两组并只让起点自由。其五，提取诱发遗忘对应判定这一动作本身的抑制作用，因此本文预测**判定所覆盖内容与其同类未覆盖内容之间**存在系统性此消彼长，可在同一份卷子的覆盖与未覆盖题目间直接测。其六，整合作为边界条件，对应本文的处置方向：让被判定内容之间形成结构，可以取消判定的抑制作用。其七，可观测性对应"进不了评价的状态进不了教学决策"，因此本文预测增加评价维度并不改善结果——只要新增维度仍在同一窗口内采样。其八，采样定理对应判定频次与重组频率的关系，因此本文预测提高判定频次不能解决问题，只会把混叠做得更精细。其九，抗混叠滤波对应"判定时写明哪些成分不予判定"，这是本文与"多测"和"晚测"都不同的第三条处方。其十，不可控不可观子空间的处置方式（证明稳定即不管），对应课堂对"这批学生就这样了"的处置，因此本文预测该处置的正当性完全依赖于观测窗口，换一个窗口即失效。

第三、第五与第九项是本文最依赖的三项。删掉第三项，本文的对象退回"遗忘"；删掉第五项，判定的主动作用消失；删掉第九项，处方退回"延迟测验"。

十二、一个班走完：从判定到配置到锁死

取初中数学"分数除法"这个单元，跟一个班走一遍。

第三周作出判定。单元测，班上二十六人不及格。这二十六人里，按本文的亚型分，至少有三类混在一起：一类正在把"除以一个数等于乘它的倒数"从口诀重述为"一个量里有几个另一个量"（表征重组，以周计）；一类已经会两种做法而不知何时用哪种（策略选择，以周计）；一类只是没做够题（重组已完成，纯熟练度问题，以天计）。**卷面上三类完全一样，都是错。**

第四周据判定配置条件。二十六人进补救组，补救内容是同型题的分步练习。这个安排对第三类完全正确——他们需要的就是多做题，两周后追上。对前两类是灾难：分步练习把"一个量里有几个另一个量"这个正在成形的重述**替换成了一套操作步骤**，因为分步练习的每一步都是操作，没有一步是关系。第一类学生的重述在这里停住，因为它不再有用武之地。

第六到第十二周，冻结构型在新条件下继续。补救组的任务由判定配置，而判定按操作正确率读数，因此补救组的表现稳定上升——他们确实会越来越会按步骤做。**判定的自我实现在这里不是比喻：**读数上升，正是因为读数所测的正是补救任务所训练的那一样。第一类学生此后遇到需要用关系而不是步骤的题，仍然做不出，而这被记作"基础太差、补了也就这样"。

如果第五周做过一次复测。同一份卷子、四周间隔、期间不教——按本文预测，第一类学生里会有相当一部分翻成"会"，因为重述在这四周里仍在无教学地进行。这一支翻转在现行安排里不可能被观察到，因为第四周就已经配置了补救，四周之后他们做的已经不是原来那份卷子。**只要配置一开始，那个反事实就永远拿不到了。**

成本对比。复测的成本是一份现成卷子加一节课的时间；配置的成本是八周的补救课时。而现行安排把便宜的那件省掉，把贵的那件做足。这不是谁的疏忽——**教务系统里有"补救组"这个字段，没有"复测结果"这个字段**，因此前者是一个可以走的流程，后者是一件需要有人特意去做的事。

这个推演出三条可以被推翻的具体主张。其一，同一次不及格里混着重组时长相差一个数量级的几类学生，而判定按同一套规则处置他们。其二，补救内容的性质会决定哪一类被救、哪一类被冻——分步练习救第三类、冻前两类，这一条可以在补救材料上直接看出来。其三，复测的成本远低于补救的成本，而前者从不做、后者一直做，原因是记录结构而非教学理念。三条都可以在任何一个初中数学教研组用现成材料查证。

十三、三家为何各自到不了

三家没有到达这个位置，不是因为疏忽。每一家各自的问题结构，恰好使这一步成为不必要的。

物理学到不了，因为它的体系里没有“据读数配置后续条件”这一步。物理学家读出一块玻璃处于玻璃态，不会因此改变这块玻璃往后能接触什么环境；读数是关于系统的一个描述，它不参与系统此后的处境。物理学确实有观测扰动系统的问题，但那是测量动作本身的能量交换，与本文所说的完全不同：本文说的是**观测的结论**改变了系统此后的输入。这一步在物理学里没有对应物，因为物质不会因为被判定为某种物态而被安排进不同的房间。

心理学到不了，因为它的方法学在结构上排除了本文的对象。提取诱发遗忘的标准范式在二十分钟到一小时内完成，这是必须的——时窗一长，日常生活的干扰就无法控制，效应会被淹没。而本文关心的重组以周或月计。一个以月计的过程在一个以分钟计的范式里根本不可能被观察到，这不是这一支不够努力，是它为了获得因果控制所必须付出的代价。它掌握的关于抑制的一切知识都是在窗内取得的，而窗外的事它在设计上就看不见。

工程学到不了，因为它的系统没有自己的往后。不可观测子空间证明稳定即可不管，这在机器上是对的：一个不可观测且稳定的状态不会因为被这样处置而变成另一个样子。而在人身上，“不可观测且被判为稳定”这件事本身会通过分层、补救、免修、放弃改变这个人往后能接触到什么，于是他确实变成了另一个样子。工程学没有“被判定者”这个角色——它的对象不会因为被如此判定而改变自己此后的轨迹，因此它从来不需要问一次判定做了什么。

三家的盲区各不相同，交集恰好是本文的位置：**一个读数在原理上无效、而据它配置的条件使被判定者只能沿这个读数继续。**物理学有读数的无效性而没有配置这一步；工程学有配置这一步而对象没有往后；心理学有对象的往后而看不见这个时间尺度。三条各占一半，合起来才是那一格。

十四、为什么常见的改革处理不了这一件

近三十年里，提高反馈频次、引入形成性评价、推行掌握学习、发展自适应测评、增加过程性评价，每一样都被推行过。本文给出的解释是形式上的：**这五类改革改的全是判定的形式与频次，而窗比与“判定是否配置后续条件”一个都没动。**

提高反馈频次改的是采样率，而采样率提不上去。课时、疲劳与教学时间都不允许把判定频次提高到重组频率的两倍以上，因此提高频次只是把混叠做得更精细——得到一条更平滑的伪曲线。这一条给出一个可查的预测：**判定频次与迁移表现之间应当存在拐点，超过它之后相关转负。**

形成性评价改的是判定的用途，不改判定的时点。它要求证据用于调整教学，而调整所依据的证据仍然取自同一个过短的窗口。用一个无效读数去调整教学，调整得越及时，冻结得越彻底。

掌握学习改的是通过的门槛，且把窗比推得更小。不达标就重学、重测，直到通过——这意味着同一个人同一内容上被反复短窗判定，被冻在更多个点上。本文预测严格执行掌握学习的班级双向翻转率反而更高，这与该支的自我评价相反。

自适应测评改的是题目的选择，不改时间尺度。算法根据作答实时调整难度，采样在毫秒级进行——而重组在周与月的尺度上。采样再密，密的也是同一个窗口内的密，跨不过那个尺度。

过程性评价改的是记录的丰富度，不改配置的依据。记录更多的过程数据本身不产生任何损害；损害来自这些数据被用来配置后续条件的那一步，而这一步在过程性评价的设计里从未被单独讨论——它被默认为过程数据的正当用途。

五类改革共享一个特征：**它们都把"判定得更好"当作目标，而没有一样把"判定与后果之间那条线"当作可以拆开的东西。**试点期间，研究者在场、样本量小、后续配置往往被暂缓，损害被临时抑制；推广之后配置自动恢复，此前的改进同时失效。

十五、与最接近的几种说法的分界

遗忘曲线与保持研究说到的是：不复习则表现随时间下降。分离线在**方向**：遗忘是单向的，本文预测双向。判定形式：若两次测量之间的翻转全部是"会到不会"，则本判断错而遗忘足够；若"不会到会"这一支在无教学条件下显著非零，则反之。这一条最便宜，且能一次分开两者。

间隔效应与延迟测验（Soderstrom & Bjork, 2015）说到的是：练习分散、测验延迟有利于长期保持。分离线在**被安排的东西**：那一支安排的是**练习**的时点，本文说的是**判定**的时点。判定形式：把练习完全等距固定，只改判定时点与判定后的配置，间隔效应预测无差异，本文预测有差异。

经典测量理论中的重测信度说到的是：同一测验两次施测的相关系数反映测量的稳定性。分离线在**真值**：那一支假定真值在两次施测之间不变，把差异全部归给误差；本文说真值在移动。判定形式：若低重测信度的项目在延长间隔后信度继续下降而**均值不移动**，则是误差；若均值系统性移动且移动方向与初次判定相关，则是位移。

掌握学习（Bloom, 1968）说到的是：不达标就重学，直到达标。分离线在**达标判定的时点**：掌握学习把判定放在教学之后立即作出，反复判定直到通过，因此它的窗比在结构上最小。判定形式：若严格执行掌握学习的班级其双向翻转率低于常规班级，则本判断错；本文预测它反而更高——因为反复的短窗判定把每个人冻在更多个点上。**这是一条相反评价，也是本文最锋利的一条分离线。**

学习率的稳定性（Koedinger, Carvalho, Liu & McLaughlin, 2023）说到的是：在大规模数据中，学习者的学习率惊人地一致，差别主要在起点。分离线在**可观测段**：那一支测的是窗内可被记录的那一段练习—表现关系。判定形式：若在该数据集中把练习序列按"是否伴随一次配置后续条件的判定"分层，两层的起点差异不变，则本文无增量；若被判定过的那一层起点系统性更低而斜率相同，则本文所指的正是"起点差异"的一个来源。**这一条可以在已公开的数据上重新分析，不需要新采集。**

最近发展区说到的是：存在一个凭借帮助可及、独立不可及的能力带。分离线在**维度**：那一支说的是难度的边界，本文说的是读数的有效期。判定形式：把难度完全固定，只改判定时点——最近发展区预测无差异，本文预测有差异。

德博拉数与流变学是形式层的上游占位者，来自物理学之外的工程学科。把本文的命题剥掉学科名词，剩下的是：**以低于对象变化速率的频率观测，得到的不是缺席，而是一个伪像。**这个结构在流变学里叫德博拉数（Reiner, 1964），在信号处理里叫混叠（Shannon, 1949），在统计学里叫时间聚合偏误，在生

态学里叫**粒度与幅度**（Wiens, 1989）——同一片林地，用不同的取样粒度会得到不同的物种—面积关系，而它们不是谁更准，是不同的量。这几位必须点名，且本文的分离线只有一条：**它们讲观测尺度决定所观测对象的类别，本文讲观测之后据此配置的条件参与了对象此后的演化**。判定形式：若判定不配置后续条件时长期表现不受损（左下格），则本文相对这些占位者有增量，因为它们只管读数不管后果；若不作配置的短窗判定同样有害，则本文可被尺度问题吸收。

成功陷阱这一类反转模板不适用于本文，必须写明。本文的形式不是“越成功越失败”，是“观测尺度决定所读到的类别”。反转模板要求存在一项被推到极端的能力，而窗比过小是课表的结构性后果，与课程不成功无关：一门长期失败的课程窗比同样小。这一条排除是必须写的，因为形式层最容易被误认到这个模板上。

划完之后再取一次交集。逐条问这几位各自结构性地看不见什么：遗忘研究把变化的方向当作给定，因此看不见反向翻转；重测信度把真值的稳定当作给定，因此看不见位移；掌握学习把“达标”当作可在教学结束时判定的事实，因此看不见判定本身的时点问题；最近发展区把能力的边界当作在任一时刻可读的，因此看不见有效期。这几行背后是同一件被当作给定的东西：

> **一个人在某一时刻“会不会某件事”，是一个有确定值的事实。**

所有派别都在争“怎么测得更准”，没有一派把“这个值在被测量的那段时间里是否稳定”当作变量——因为一旦承认它在动、且动得慢于测量，“**测得准**”这件事本身就失去了对象。这是本文动摇的那块地。

十六、一处被逼出来的自相矛盾

上一节论证本文不能被既有说法还原。本节论证一件更强的事：教育测量这一行，在不承认本文变量的前提下无法自治。

这一行同时持有两条命题。**甲**：信度估计要求真分数在测量期间稳定——这不是它的推论，是重测信度这个量得以被定义的前提。**乙**：教育测验测的是学习，而学习是变化。

两条在同一套操作里互相拆台。若被测者在测量窗口内仍在变化，重测信度低就有两个不可分的来源：测量误差与真实变化。这一行现有的处置是**把重测间隔取短**，以“控制真实变化”——而按本文，缩短间隔正是把窗比推小，使判定落在最不该判定的时点上。于是：**为了让甲成立所做的事，恰好使乙所测的那个东西被冻住。**

这一行并非没有察觉。它的应对是引入“稳定性系数”与“等值性系数”的区分，并在手册中提醒间隔的选择须权衡记忆效应与真实变化。但这条出路是循环的：要判断两次测量之间的差异有多少来自真实变化，需要一个独立于测验的变化速率估计；而这一行获得变化速率估计的唯一途径，正是重复施测。**要让甲成立必须假定变化很慢，而变化是否很慢正是待检验的。**

因此这一行面临二选一：承认真值在测量期间移动，则信度的经典定义需要重写为一个带时间尺度的量；或坚持真值稳定，则必须解释为何“不会到会”的无教学翻转会发生——而这一支自己的重测数据里就有这些案例，它们通常被作为异常值处理。

本节的主张有边界，也留了退路：若在实测中“不会到会”的无教学翻转率低到可忽略（比如低于百分之三），则甲乙的冲突近似不存在，本节失败。这一条可以用现成的重测数据直接查。

十七、与同批另两篇的分工

与《先挡住其余的》。那一篇处理的是为使一处改变而作的悬置在悬置期内结成的连接，读数是悬置密度与脱悬率。分离线在位置：那一篇管一次改变的伴生物，本文管它的出口。判定形式：一个把所有判定都推迟到重组完成之后的班级，本文预测其读数有效，那一篇预测若悬置从未拆解，推迟判定只是更晚地读到同一批连接的后果。两个预测同时成立且不冲突。

与《开着的时候没人来》。那一篇处理的是上一轮打开的感受窗口与本轮内容不对齐，读数是对窗率。分离线在轮次关系：那一篇管入口，本文管出口。判定形式：一个对窗率极高、每一轮都精确接上上一轮涌现的班级，若它同时每周作一次配置后续条件的判定，那一篇判它健康，本文判它落在靶格——因为对上窗之后启动的那次重组，恰恰是最慢的那一类，也最容易被下一次判定冻住。这一条是三篇里最反直觉的一处：对窗做得越好，被冻住的东西越要紧。

三篇合起来才说得出一句：同一次改变有入口、伴生物与出口三处可以失效；三处的读数数学上独立、可同批测量，而效应相乘不相加。本文单独看会得出“延迟判定”这个处方，而那一篇会指出延迟判定并不拆除已结成的连接，另一篇会指出延迟判定不改变窗口的开合——三处都对上，才有一次完整的改变。

十八、五类场景中的窗比，与其中两处收窄

中学数学的单元测是窗比远小于一的标准形态。教学到测验通常隔一到两周，而分数、比例、函数这几类概念的重组在文献里以月为量级。这一格的双向翻转率应当最高，而它几乎从不被测——因为单元测的卷子讲评之后就作废了，没有人会隔四周原样再发一次。

编程训练营的周考是方差最大的一格。递归、指针、异步这几个概念的重组时长个体差异极大，从三天到三个月都有。因此同一次周考里，一部分人的窗比大于一、一部分人远小于一，而判定对两部分人同等对待。本文预测这一格的双向翻转率最高且分布双峰。这一格也最容易做实验，因为训练营的课程周期短、班级多、数据全在平台上。

医学住院医的当场技能评估是窗比大于一的一格。缝合、穿刺、插管这类单项操作的重组以次计，当场判定是有效的。第一处收窄在这里：本文对这一格不适用，且必须明说——把本文的处方（明写不予判定）用到这一格，会把本来可以当场判定的东西也悬起来，反而降低训练效率。窗比大的场合，及时判定是对的。

语言能力分级是落差可以直接算出来的一格。欧洲语言共同参考框架的配套材料公开给出了相邻级别之间的指导学时，量级在数百小时；而多数语言课程按学期评级，一个学期的课时通常远低于这个数。**这一格的窗比可以从公开文件里直接算，而且算出来的落差不是任何一所学校的执行问题**——它是课程周期与能力跃迁尺度之间的结构性差距。

工业技能认证是窗比极大的一格：单项操作以天计，认证一次通过后长期有效。这一格反过来提示，**窗比过大同样有代价**——认证之后不再复测，被认证者此后的老化无人知晓，而这正是物理学那一处反向约束所说的上界。

第二处收窄由这五格合起来逼出：本文的处方不是“判定越晚越好”，也不是“判定越少越好”，而是**窗比应当被知道**。知道之后怎么办因格而异：窗比远小于一的格子里，处置是把配置与判定解耦；窗比大于一的格子里，处置是照常判定；窗比极大的格子里，处置是恢复复测。本文给不出一个统一的应然数值，这是它的限度，也是它相对于“多测”与“晚测”这两条口号的全部差别。

十九、三个领域反过来加上的限制

物理学削掉了本文最自然的一条处方。若只看德博拉数，本文会写下一句更强的话：把观测窗口拉长到超过重组时长即可。而物理老化证明，等得越久，系统本身已经因等待而改变——测到的已不是当初那个系统。翻回课堂：**判定间隔存在上界**，推迟到某个程度之后，测的是“这段时间里他自己长成的样子”，而不是那次教学的效果。这一处削掉的是“越晚越好”。

工程学削掉了本文的另一条自然处方，并改变了处方的形态。采样定理要求采样率高于最高频率的两倍；课堂无法提高判定频次——课时、疲劳与教学时间都不允许，而且提高频次会加重抑制。工程的现实对策不是提高采样率，是**抗混叠滤波**：在采样之前把快于半采样率的成分滤掉，宁可明写“这部分我不测”，也不让它伪装成慢变化。翻回课堂，处方因此从“多测”或“晚测”变成第三种：**判定时明写哪些成分不予判定**，并在记录里为它们留一个“本次不判”的位置，而不是默认为不会。这一处改的是处方的形态，不是它的强度。

心理学削掉了本文的适用范围。提取诱发遗忘说明，不判定也不是中性的——一次提取就会抑制同类竞争项，而任何形式的练习都是提取。因此本文不能主张“少判定”就更好；它只能主张：**判定之后据此配置条件这一步，是可以被单独改动的那一处**。适用范围因此收窄到“判定会配置后续条件”的场合，对不配置条件的诊断性活动，本文无话可说。

二十、不适用的边界

三处反向约束已经削掉了本文的价值排序与部分处方。这一节把边界写全，共五条。

其一，判定不配置后续条件的场合不适用。这是辨别格的左下角：一次不计分、不分层、不改变任何安排的诊断性活动，即使窗比很小也不产生冻结构型。本文对它无话可说，而这一格在实践中并不罕见——它正是"低风险形成性评价"的定义。

其二，重组时长短于观测窗口的内容不适用。熟练化、程序性技能、事实性记忆，这几类的重组以次或以天计，两周的判定间隔对它们绰绰有余。把这类内容算进窗比是记错了账，而它们在课堂总量里占的比重不小。

其三，安全关键的即时判定不适用。实验室操作、临床技能、机械操作、应急程序中的合格判定必须当场作出，此处不存在推迟的余地。这些场合的处置是在模拟环境中安排复测，而不是延后真实判定；把本文的处方用在这里是危险的。

其四，一次性的资格判定不适用。当一次判定之后不再有可配置的后续条件（结业、颁证、终局性考试），冻结构型的第三个必要条件不满足。它可能仍然不公平，但那属于测量公平的问题，不属于本文。

其五，群体层面的统计判定不适用。本文的对象是个体身上的构型，而群体均值在窗比小的条件下反而可能稳定——因为不同个体的重组处在不同阶段，均值把它们抹平了。用班级平均分来检验本文的机制会系统性地得到零结果，而那不是本文被证伪，是分析单位选错了。

五条合起来把射程收到：**有连续轮次、判定会配置后续条件、内容的重组时长长于判定间隔、且分析单位是个体的那一类教学。**这个范围小于课堂判定总量的一半。

二十一、经验研究设计

本文的三个量都必须能在既有课堂里测出来。

核心设计是双时点复测。同一份判定卷，在原时点与四周后各做一次，中间不教学。"不教学"必须有可核证据，且须三源一致：教案版本、平台日志、作业记录。三源缺一，该样本作废——这一条不能松，因为本文最要紧的那一支（不会到会）恰恰最容易被中间偷偷讲过污染。

四类条件把差别收窄到一个参数上。**A 组**常规判定并据以配置后续条件。**B 组**常规判定但不配置——判定照做、结果照记，而分层、补救、跳过一律不依据它。**C 组**把判定推迟到估计的重组完成之后，并据以配置。**D 组**在原时点判定，但对识别为仍在重组的成分明写"本次不判定"，其余照常配置。

三处对照各回答一个问题。**A 与 B** 检验"配置"这一步是不是损害的来源，这是本文靶格设定的直接检验。**A 与 C** 检验单纯延迟够不够。**C 与 D** 是本文最要紧的一处——工程学那一处反向约束预测 D 优于 C，即抗混叠式的处置优于降低采样率；若实测 C 优于 D，那一处反向约束不成立。

重组时长的估计规程必须预注册，否则它会被事后调成想要的数。做法是在 **B 组**（判定不配置）中跟踪同一批学习者的表现随时间变化，取达到平台期所需时长的下四分位作为下界，并在主分析中报告该值取不同数时结论是否稳健。**这个估计有一处循环风险**：用一个理论所预言的现象去估该理论的参数。本文不掩饰它，只能靠预注册与敏感性分析降低，不能消除。

双向翻转率的编码要求分方向报告，且"不会到会"这一支须逐例核证期间无教学。编码者不得看该学生的原判定结果之外的任何材料。按亚型分层是强制的——不分层则两个方向互相抵消。

数据分四层：基线知识与先前成绩；判定时点与配置动作的记录（教案与教务系统里现成）；两次判定结果；十二周后的无支持迁移。

最低成本的先导研究不需要任何新采集。在已公开的大规模学习平台数据中，按"该练习序列是否伴随一次配置后续条件的判定"分层，比较两层学习曲线的起点与斜率。若被判定过的那一层起点系统性更低而斜率相同，本文所指的正是"起点差异"的一个来源。这一条完全在既有数据上完成，一个月之内可以跑完。

统计上以判定事件为单位，嵌套于学生、嵌套于班级。核心检验是"不会到会"支的比例与窗比之间的剂量—反应，而不是组均值的比较。

二十二、证伪条件与轮次预测

最强的竞争解释是一句朴素的话：其实不就是学生忘了、或者本来就没学扎实吗。

排除它需要一个只有本机制才预测得到的变量。本文取的是**双向翻转中"不会到会"那一支**：在两次测量之间无任何教学的条件下，遗忘解释在原理上预测不出这一支，学得扎不扎实同样预测不出。因此本文的核心证伪条件是：

＞ **控制总练习量、内容难度、教师效应与先前成绩后，若"不会到会"的无教学翻转率不显著大于零，或它与窗比之间不存在剂量—反应关系，则本判断为伪。**

据此写出七条。

其一，如上；这是核心条。

其二，若窗比小而判定不配置后续条件的班级（左下格）在长期表现上与配置条件的班级无差异，则靶格设定错误，第二根轴应当换掉。

其三，若延迟判定与"判定时写明哪些成分不予判定"两种处置在长期表现上等效，则工程学那一处的处方形态之别没有经验意义。

其四，若复测差异随间隔呈线性而非幂律，则物理老化那一条的形式对应不成立，本文只剩比喻。

其五，若判定所覆盖内容与其同类未覆盖内容之间不存在系统性此消彼长，则提取抑制那一处的兑现失败。

其六，若严格执行掌握学习的班级其双向翻转率低于常规班级，则本文与掌握学习的相反评价被推翻。

其七，若在既有大规模学习数据中，按"是否伴随一次配置后续条件的判定"分层后两层起点无差异，则本文对学习率研究无增量。

第一条与第七条最便宜。第一条只需要一份现成卷子、四周间隔与不教学的约定；第七条完全不需要新采集，只需要在已公开的数据上重新分层。

轮次预测。本文的动力主张不是相关，是一个有次序的循环。

> 第 t 轮作出判定 → 第 t+1 轮据判定配置条件（分层、补救、跳过）→ 第 t+2 轮被冻住者的慢速演化在新条件下沿冻结构型继续，因而其表现与判定一致 → 第 t+3 轮的发起处从"内容"转到"这批学生的能力"，处置是提高判定频次，窗比进一步变小。

四步中任何一步顺序系统性相反，模型即需修正：若能力归因先于判定出现，则判定是归因的结果而非原因；若配置条件的改变与判定结果无关，则第二步不成立。

二十三、判定可以怎么改：三个动作

本文首先是一项解释，不是评价方案。但一旦把冻结构型看见，判定的做法会变，且变的地方比想象中便宜。

第一，给"本次不判定"一个字段，并按实际填写率验收。成本最低的一件：在成绩记录里增加一个既非"会"也非"不会"的选项，其含义是"该成分仍在重组，本次不作判定"。这不需要任何新技术，只需要教务系统加一列。要紧的是验收方式——一个填写率长期为零的字段与没有这个字段是同一格，而且更有害，因为它已经把合规账结清了。

第二，复测一次，并分方向报告。一份现成的卷子、四周的间隔、期间不教学的约定，就是本文全部主张的最便宜的一次检验。分方向报告是关键：只报总翻转率，两个方向会互相抵消。这一条不需要任何人相信本文，做一次就知道。

第三，把配置与判定解耦。判定照做——它有别的用处，教师需要它来安排讲评。而分层、补救、免修、放弃这几项配置动作推迟到复测之后。这一条改的不是判定，是**判定与后果之间那条线**，而那条线在多数学校里是自动连上的，从未被人当作一个可以拆开的设计选择。

三条里只有第三条涉及制度安排，前两条是记录方式的改变。这一点本身可检验：若在完全不改变配置流程的条件下，仅按前两条运行的班级与对照班无差异，本文的处方失效。

二十四、反噬预言与伦理限制

窗比与双向翻转率一旦进入考核，最省事的达标办法会摧毁它们。

第一种是**推迟判定而不改任何别的**。窗比在账面上立刻变大，而判定仍然一次性、仍然配置后续条件、仍然把仍在进行的重组冻住——只是冻得晚一点。

第二种是**把复测做成走过场**：同一份卷子隔四周再发一次，而中间悄悄讲过，或者复测不计分因而学生不认真作答。前者使翻转率虚低，后者使它虚高，两者都无法从数据上分辨。

第三种最难查：**把"本次不判定"这一栏做成一个从不被填写的字段**。它挂在评价表上、写进制度文件、计入合规检查，而默认路径仍然是每一项都给出会与不会。

对此本文看不到彻底的出口。能规定的只有一条判定纪律：**复测的有效性以"期间无教学"的可核证据为准——教案版本、平台日志与作业记录三者一致才算数**；以及一条验收纪律：**"本次不判定"这一栏的验收看实际填写率，功能存在而填写率为零，与没有这一栏是同一格**。

因为窗比与翻转率都是可以拿去考核人的比率，伦理限制必须写死三条。

第一，这两个读数指的是**判定的时点**，不是学生的能力，也不是教师的水平。用它给学生贴"慢"或给教师排名，是对本文的误用。

第二，**不得为了把窗比做好看而在安全关键教学中推迟必须立即作出的判定**。实验室操作、临床技能与应急程序中的合格判定必须当场作出，此处的处置是在模拟环境中安排复测，而不是延后真实判定。

第三，若学校已据这两个读数作出不利安排（分层、免修、劝退），**必须公开编码规则、原始证据与复核通道**，并允许被评价者要求一次符合本文条件的复测。

二十五、自反

按本文自己的判据，这篇文章的窗比小于一。

它对一个需要数年才能重组的问题，在一次写作里作出了判定，并把这个判定以完成态交出。它的读者也将在一次阅读里对它作出判定，而这次判定同样会配置后续条件——引用、不引用、纳入综述的哪一格。本文因此正落在自己的靶格里，且落得比多数文本更深：它连一次复测的可能都没有留下，因为一篇发表出去的文章不会在四周后原样再来一遍。

本文的双向翻转率无法自测。**能说的只有一句：本文把可以让自己失败的那七条写得比结论更具体，其中两条不需要任何新数据**。这是本文能为自己配的唯一一件东西——它不能延长自己的窗口，只能把复测的工具交出去。

二十六、分层引用，与三个本文解释不了的洞

本文可以被分层引用。**第一层**是"冻结构型"作为一类对象，最强也最易倒——只要证明无教学期间不发生变化为正的变化，这一层就没有了。**第二层**是那张二维辨别格，即使命名不成立，把窗比与"是否配置后续条件"分成两根轴仍是可用的诊断工具。**第三层**是一条不依赖前两层的经验主张：同一份判定在无教学的两个时点之间会双向翻转，且"不会到会"这一支非零。第三层最稳，也最便宜。

有三个洞本文解释不了。

第一个洞：本文测不准重组时长，而窗比的分母正是它。估计规程有循环风险，敏感性分析只能降低不能消除。这使得窗比在实践中更像一个定性判断而非一个可比的数，而本文的整套论证却把它当作一个比值来用。双向翻转率不依赖分母，因此本文实际能交付的读数只有一个。

第二个洞：本文说不出被冻住之后的演化能不能被逆转。物理老化在材料学里可以被回火消除——重新加热到玻璃转变温度以上再缓慢冷却。人身上有没有对应物，本文不知道，也不敢类推。若存在，本文关于“配置条件使人只能沿冻结结构型继续”的整个主张需要加上一个时限。

第三个洞：本文无法分辨“判定造成的冻结”与“判定只是恰好发生在那一刻”。两者在数据上都表现为翻转与窗比相关。要分开它们，需要一个判定时点被随机分配的设计，而在真实学校里随机分配判定时点几乎不可能获得伦理批准——因为它意味着一部分学生的分层依据被人为推迟或提前。这是本文最实质的一处方法学困难，本文没有解决办法。

二十七、三条本文预测而尚无人测的现象

一个判断若只解释已知，仍可能是旧说法的新名字。本文因此列出三条它预测、而据本文所知尚无人测过的现象。它们不是推论的装饰，每一条不成立都要动本文的一部分。

其一，不及格者的复测翻转率应当高于及格者。通常的预期相反——及格者被认为掌握得更稳，因此更不容易翻转。而按本文，及格意味着重组已经完成或接近完成，翻转空间小；不及格者中混着大量重组进行到一半的人，翻转空间大。因此本文预测**双向翻转率与原判定结果之间存在负相关**，且这一相关在按亚型分层后主要来自表征重组与整合两类。这一条只需一份现成卷子做两次即可查，而它与直觉相反，因此是一个便宜的风险点。

其二，补救的内容性质应当决定翻转的方向。以操作步骤为主的补救材料会把仍在重述的学习者冻住，因而其组内“不会到会”的自发翻转率应当低于不作任何补救的对照组——**补救比不补救更糟，在这一格里**。这一条的实验设计很直接：不及格者随机分为分步补救、关系型补救与不补救三组，四周后复测。若不补救组的自发翻转率不高于分步补救组，本条为伪。

其三，“本次不判定”字段的填写率与教师的学科知识相关，而与教学理念无关。要认出一个学生正在重组而不是没学会，需要认出他那半套重述是什么——这依赖学科知识，不依赖对学生的态度。因此本文预测该字段的填写率与教师的学科背景相关，而与教学信念问卷得分无关。这一条若不成立，说明识别“仍在重组”不需要学科知识，那么本文关于该字段可行性的整套设想需要重写——它将变成一件任何人都能做、也因此任何人都会做成走过场的事。

三条都不需要新的测量工具，只需要有人把已经在做的事按不同的方式记一次。

二十八、读数的可比性：三处本文自己交代不清的地方

一个读数若不能跨课堂、跨学科比较，它就只是某个人的记录习惯。本文的两个读数在可比性上各有难处，必须写在正文里而不是脚注里。

窗比的分母不可跨学科比较。分数概念的重组时长与递归概念的重组时长不是同一件事的两个数值，它们的"重组"指的可能是不同类型的过程。因此窗比只能在同一学科、同一内容类型内部比较；跨学科的窗比排名没有意义，而这恰恰是最容易被做出来的那种排名。本文对此的处置是：**窗比只用于同一门课的纵向比较，不用于横向排序**，并把这一条写进伦理限制。

双向翻转率的分母依赖判定的难度设置。一份很难的卷子会把大量已完成重组的人也判为不会，从而抬高"不会到会"这一支；一份很容易的卷子则相反。因此翻转率必须与判定的通过率一同报告，单独报告的翻转率不可比。这一条使本文最便宜的那件检验多了一个约束：复测必须用**同一份卷子**，而不是同等难度的另一份——后者引入的差异无法与真实翻转分开。

亚型分层依赖学科知识，因而不可自动化。要判断一个学生正处在表征重组还是策略选择，需要认出他那半套重述是什么，这依赖学科知识。任何试图用作答数据自动分层的做法，都只能按对错模式聚类，而对错模式与重组类型之间没有稳定映射——第一类与第三类的错误在卷面上可以完全相同。这意味着本文的读数在可预见的将来**不能被平台自动产出**，只能由懂这门学科的人手工编码。这一条大幅提高了本文的实施成本，也大幅降低了它被做成考核指标的风险；两个后果一好一坏，本文如实写下，不作取舍。

三处合起来说明：本文交付的不是一个可以填进仪表盘的数，是一套需要人来做的编码。这是它的限度，也可能是它唯一的保护。

二十九、一处必须说清的伦理前提：谁有权要求复测

本文交出的最便宜那件检验——同一份判定隔四周原样再做一次——在制度上并不中性，因此必须把它的伦理前提说清，否则它会以最坏的方式被实施。

复测不能是一次新的考试。若复测计入成绩、影响排名、决定分组，它就变成了第二次判定，而按本文，第二次短窗判定只是把人冻在另一个点上。复测的唯一正当用途是**核验第一次判定的有效性**，因此它必须不计分、不排名、不进档案。这一条若守不住，本文的处方比不实施更糟。

要求复测的权利应当在被判定者一侧。一个学生若认为那次判定作得太早，他应当可以要求一次符合条件的复测；而这个权利在现行制度里不存在——申诉的对象是判分的对错，不是判定的时点。本文因此主张一条具体的制度改动：**在任何据判定作出的不利安排（分层、免修、劝退）生效之前，被安排者有一次要求复测的机会**，且复测须满足期间无教学的核证条件。这条改动的成本是一份现成卷子与四周的等待。

复测的结果不得用于追究教师。若"判定被复测推翻"成为教师的失分项，教师的理性反应是把第一次判定作得更保守、更晚、更少——而那会把本文的第三个动作（配置与判定解耦）挤掉，因为解耦要求判定照常作出。读数指向的是**判定的时点与配置之间那条线**，不是任何人的水平。

三条合起来说明一件事：本文所主张的那一件，其可行性完全依赖于复测能否保持低风险。一旦它被卷入问责，它会立刻变成它所批评的那个东西。本文看不到防止这一点的技术办法，只能把它写在这里。

三十、结论与一条写死日期的赌注

课堂上学习很少真正到位，不是因为它没有在进行。它在进行，只是进行得慢于课表。课堂每隔一到两周作一次判定，而许多内容的重组以月计；于是判定系统性地把仍在进行的变化读作零，并据此配置此后的一切条件，使被判定者只能从被冻住的那个构型继续。

三个领域各出一份不可替代的力。物理学给出两条最承重的形式约束——**"有没有在变"是关于观测窗口的读数，不是关于系统的属性；以及被判为不变的系统仍在演化，且以被判定那一刻的构型为起点。**心理学说明判定本身不是中性的观测，它会主动抑制没有被它覆盖的那些。工程学则给出了处方的形态：不是提高采样率，是在采样之前明写哪些成分不予判定。

这一理论把课堂评价的问题从"测得准不准"改为"测的时候它停下来了没有"；把改革方向从"更频繁的反馈"改为"为不予判定留一个位置"；把归因从"这批学生基础差"改为"这批学生被冻在哪一步"。

本文给出一条写死日期的赌注。**到 2031 年 6 月 30 日之前**，若至少三项在同一制度环境内、不少于八个平行单元、控制总练习量、内容难度、教师效应与先前成绩之后开展的研究报告：同一份判定在无教学的两个时点之间发生双向翻转，其中"不会到会"这一支显著非零，且该支的比例与窗比之间存在剂量—反应关系，则本文的核心判断获得初步支持。

以下明确**不算命中**：仅报告重测信度系数；仅报告遗忘曲线或保持率；仅延长测验间隔而未核证期间无教学；仅比较两个地区或两所学校；把延迟判定与补充教学、补救辅导打包实施而无法分离；把"本次不判定"栏的存在计为已实施而不查填写率。若符合上述条件的多项研究均未发现"不会到会"这一支，本文应被放弃或降格为一个描述性提法。

课堂不需要放弃判定——不判就无法安排。它需要的只是：在按判定安排一个人的后半年之前，先问一句那次判定是在他停下来之后作的，还是在他还动着的时候作的。

注释

一、本文所用"弛豫时长""德博拉数""玻璃转变""物理老化""可观测性""混叠""抗混叠"等术语取各自领域的通行含义，用于提供关于观测窗口与被观测对象关系的形式约束，不作为关于神经过程的经验主张。

二、第十二节的自相矛盾主张限于"重测信度的定义前提"与"教育测验测的是变化"两条之间的取舍，不对教育测量学的全部工作；该行关于稳定性系数与等值性系数的区分正是这种意识的产物。

三、本文为理论性文章，不报告作者新采集的实验数据。文中的经验命题、指标与研究方案均需通过预注册研究与独立复核检验。

参考文献

- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087. doi:10.1037/0278-7393.20.5.1063
- Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 608–629. doi:10.1037/0278-7393.25.3.608
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1–12.
- Debenedetti, P. G., & Stillinger, F. H. (2001). Supercooled liquids and the glass transition. *Nature*, 410, 259–267. doi:10.1038/35065704
- Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 1(2), 152–192. doi:10.1137/0301010
- Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, 120(13), e2221311120. doi:10.1073/pnas.2221311120
- Reiner, M. (1964). The Deborah number. *Physics Today*, 17(1), 62. doi:10.1063/1.3051374
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10–21. doi:10.1109/JRPROC.1949.232969
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199. doi:10.1177/1745691615569000
- Struik, L. C. E. (1978). *Physical Aging in Amorphous Polymers and Other Materials*. Elsevier.
- Wiens, J. A. (1989). Spatial scaling in ecology. *Functional Ecology*, 3(4), 385–397. doi:10.2307/2389612

字数与版本：正文约 1.9 万汉字（不含参考文献）；v1.0，2026 年 8 月 3 日。

人机分工声明：研究问题与三个领域的选取由作者提出；人工智能参与文献检索、论证结构化与文档制作。文中涉及实证效果、指标阈值与研究方案的内容需在正式投稿前由作者逐项核验，并按目标期刊规范补充伦理审批与数据计划。

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