

无主时段：一个被教育测量系统折为零的时间类别

——概念界定、三条命题与基于 575 万条作答记录的预注册检验

王德生 Claude

（德麦国际出版社／SDE Universes 研究部，新加坡 048622）

中图分类号 G434 文献标识码 A

摘 要

目的 在教育测量与教育技术的既有框架中，学习者两个动作之间没有产出的时间被一致地当作待压缩的成本，其代价从未被系统检验。本研究提出并检验一个此前未被单独指认的时间类别——无主时段，即上一程序已交出控制、下一程序尚未接手、任何一方均不持有启动权的一段时间，并考察它与学习效果的关系形态。

方法 本研究以心脏电生理的逸搏间期与超速抑制、燃烧动力学的点火延迟期与爆震、微生物学的迟滞期与传代失能为参照，析出三者共享而未被教育研究检验的形式结构。据此提出抑制命题、非零下界命题与无主性守恒命题三条，并转化为四条可证伪假设。研究采用预注册的观察设计，预注册文本于任何结果被观察之前写定并计算 SHA-256 摘要。数据取自 ASSISTments 2009—2010 与 2012—2013 两个公开数据集，经预设的纳入排除标准后，分别保留 286 228 条与 5 468 190 条作答记录，涉及 26 300 名学习者。分析采用学习者内标准化的对数作答潜伏期，以混合协变量的二次逻辑回归与最小二乘回归估计。

结果 （1）抑制假设未获支持，且在大样本上方向与预测相反：外部起搏强度每提高一个单位，撤除后的自主作答潜伏期的对数降低 0.340 ($p = 8.1 \times 10^{-10}$)。（2）非零下界假设在两个数据集上一致获得支持：作答潜伏期与正确率呈倒 U 型，二次项系数分别为 -0.152 与 -0.135 (p 均小于 0.001)，极值点位于学习者自身分布内部 ($z^* = -0.200$ 与 -0.038)，显著高于第 5 百分位。（3）撤销规则未被触发。（4）第四条假设因变量间存在定义性绑定而不可解释，该缺陷由本研究自身的操作化失误造成，替代检验以探索性身份报告。

结论 “作答越快越好”这一被广泛内置于教育技术产品中的默认设定不成立；作答潜伏期与学习结果之间存在非单调关系，其极值点不在最快端。抑制命题被本研究自身的检验推翻，据此改判为：外部起搏并不延长撤除后的自启动潜伏期，而是直接压缩无主时段本身，使学习者的行为收敛为“立即作答”与“立即求助”的二值模式。本研究进一步为反馈时机之争提供了一个可检验的整合解释，并给出一项可在一个班级、两周内完成的判决性实验设计。

本研究的创新点

1. 界定了一个此前未被单独指认的时间类别——无主时段，给出三种形态、七条与近邻概念的分离条件及一份可供现场使用的清点手册；其判定不依赖当事人自述，而依赖责任归属结构。

2. 在两个学年、575 万条作答记录上以预注册方式检验了作答潜伏期与学习结果的关系形态，发现该关系为倒 U 型且极值点不在最快端，从而否证了教育技术产品中广泛内置的”作答越快越好”这一默认设定。
3. 报告了一条核心命题被本研究自身检验推翻的完整过程，并据此改判为一个更强、更易被证伪的表述；同时公开报告了一条假设因本研究操作化失误而作废的全过程。
4. 为反馈时机之争提出一个可预先计算的整合解释，并给出一项可在一个班级、两周内完成、一次判定四类近邻理论的判决性实验设计。

关键词 无主时段；启动权；作答潜伏期；非单调关系；反馈时机；预注册研究；智能辅导系统

Abstract

Purpose. Across educational measurement and educational technology, the interval between two learner actions in which no output is produced is uniformly treated as a cost to be minimized; the price of that minimization has never been systematically examined. This study proposes and tests a previously unnamed temporal category — the unclaimed interval, defined as a stretch of time in which the preceding routine has relinquished control, the succeeding routine has not yet assumed it, and no party holds the right of initiation — and examines the functional form of its relation to learning outcomes.

Method. Drawing on escape intervals and overdrive suppression in cardiac electrophysiology, ignition delay and knock in combustion science, and the lag phase and serial-passage adaptation loss in microbiology, the study extracts a formal structure shared by the three and untested in education. Three propositions are derived and translated into four falsifiable hypotheses. A preregistered observational design was employed; the preregistration was fixed and hashed (SHA-256) before any outcome was observed. Data were drawn from the public ASSISTments 2009–2010 and 2012–2013 datasets, yielding 286 228 and 5 468 190 responses from 26 300 learners after preregistered inclusion criteria.

Results. (1) The suppression hypothesis was not supported and reversed in direction in the large sample ($\beta = -0.340$, $p = 8.1 \times 10^{-10}$). (2) The non-zero-lower-bound hypothesis was supported in both datasets: response latency and accuracy form an inverted U, with quadratic terms of -0.152 and -0.135 (both $p < .001$) and vertices interior to each learner's own distribution. (3) The preregistered abandonment rule was not triggered. (4) The fourth hypothesis proved uninterpretable owing to a definitional binding between two fields, an operationalization error of this study's own making.

Conclusion. The default assumption that faster responding is better, widely embedded in educational technology, does not hold. The suppression proposition was overturned by the study's own test and is accordingly revised: external pacing does not lengthen post-

withdrawal latency but compresses the unclaimed interval itself, driving learner behaviour toward a binary pattern of immediate answering or immediate help-seeking.

Keywords. unclaimed interval; right of initiation; response latency; non-monotonic relation; feedback timing; preregistration; intelligent tutoring systems

1 问题提出

1.1 研究背景：一个未经论证的指标转向

在当前的教育技术产品中，“更快”几乎处处是正向的。作答计时条、连击奖励、限时挑战、下一题自动弹出、以秒计的自动批改与即时提示，构成了一整套指向同一方向的设计。与之配套的学习分析指标体系，也普遍将作答用时视为难度或投入度的代理变量，将其缩短视为效率改善。

这一取向的来源并非教育学，而是人机交互。自 Miller (1968) 与 Shneiderman (1984) 以来，交互系统的设计准则将系统响应时间视为越短越好的量，这一准则有充分的经验基础：人在等待系统响应的那段时间里确实无法完成任何有价值的工作。问题在于，该准则进入教育软件之后发生了一次未被论证的主语转换——从“系统应当尽快响应人”，转为“人应当尽快响应系统”。

这次转换将同一个负号加在了性质完全不同的对象上。人在题目呈现之后的那段时间里并非无所事事，而是在从事记录系统无从知晓的活动。将其一律记为成本，等价于假定该段时间不携带关于学习者的任何信息。

1.2 已有研究的三个缺口

已有研究并非没有触及无产出时间的价值。等待时间研究 (Rowe, 1974)、孵化效应研究 (Wallas, 1926; Sio & Ormerod, 2009) 与间隔效应研究 (Cepeda 等, 2006) 均在不同层面论证了“不立即推进”的益处。但这三条线索共享三个缺口。

其一，均以时长为唯一自变量。三者操纵的都是“多久”，无一操纵“这段时间归谁”。在课堂情境中，教师提问后的三秒与教师转身板书时的三秒，在时长上完全相同，而在责任归属结构上截然不同——前者中学习者欠教师一个回答，后者中无人欠任何人任何东西。已有研究不区分二者。

其二，均预设单调关系。Rowe (1974) 的结论在其研究区间内是单调的：更长更好。孵化与间隔研究虽偶有非单调的报告，但主流的操作化仍是“更长的间隔对保持更有利”。没有研究把“最优点在何处”当作一个需要估计的参数。

其三，缺少来自教育之外的形式约束。教育研究对无产出时间的处理，长期停留在“它有时有用”这一软主张上，而软主张难以被任何数据推翻。若要使之可证伪，需要引入一个能给出函数形态约束的外部参照。

1.3 三个跨领域参照的引入理由

本研究引入心脏电生理、燃烧动力学与微生物学三个参照领域。选择依据有三：

第一，三者均以精确的量刻画了一段无产出时间（逸搏间期、点火延迟期、迟滞期），且均建立了成熟的测量方法与理论分歧，具备可移植的形式结构。

第二，三者与教育研究不存在既有的理论交换。检索显示，教育研究文献中对“超速抑制”“爆震临界”“迟滞期模型”的引用接近于零，这排除了本研究只是在复述某个已被引进教育学的概念。

第三，也是最关键的一点，三者各自都在自己的成熟知识内部与“无产出时间是纯成本”这一预设冲突，并且冲突的方式互不相同——分别在**认识论**（测量必须在撤除之后）、**函数形态**（存在非零下界）与**归属**（它不属于任何一方）三个层面。这使得三者不是同一论证的三次重复，而是同一结构的三个不同侧面。

需要说明的是，本研究并非主张教育与上述三个领域在机制上同构。生理、化学与生物过程的机制不可能直接解释学习。本研究移植的是**形式结构与测量策略**，而非机制。这一限定贯穿全文，并在第 7 部分作为局限重申。

1.4 研究问题与贡献

基于上述，本研究提出三个研究问题：

RQ1 在既有的教育测量框架中被折为零的那段时间，能否被界定为一个具有明确边界、可与近邻概念区分的独立类别？

RQ2 作答潜伏期与学习结果之间的关系形态是单调的还是非单调的？若为非单调，其极值点是否位于最快端？

RQ3 外部起搏（即系统持续提供提示与推进）对学习者的撤除后的自主启动，产生何种可测量的影响？

本研究的贡献有三。其一，提出无主时段这一概念，给出四条可操作的边界与一份清点手册。其二，在两个大规模公开数据集上执行预注册检验，为作答潜伏期与正确率的非单调关系提供一致证据。其三，如实报告一条核心假设被推翻、以及一条假设因自身操作化缺陷而失效的全过程，并据此改判理论命题。

1.5 本文的结构

第 2 部分梳理等待时间、孵化、间隔与反馈时机、帮助寻求、学习者建模五类教育研究文献，以及三个跨领域参照的文献基础，并析出其共同预设。第 3 部分给出无主时段的概念界定、可操作化清点规则、与七个近邻概念的分离条件、三条命题及四条研究假设。第 4 部分说明数据、变量、分析策略与预注册程序。第 5 部分报告预注册检验结果与探索性分析。第 6 部分讨论命题改判、对反馈时机之争的理论意涵、判决性实验设计、三个参照领域的反向约束与形式移植的合法性。第 7 部分列出局限与后续研究路径。第 8 部分给出结论及对教学、技术设计、评价制度与研究方法四方面的启示。

2 文献综述

2.1 等待时间研究

Rowe (1974) 在科学课堂的观察中发现, 教师提问后的平均等待时间不足一秒; 当等待时间延长至三秒以上时, 学生应答的长度、推理复杂度、主动发言率与教师提问的认知层次均显著上升。Rowe (1986) 进一步区分了”等待时间一”(提问后至学生开始应答)与”等待时间二”(学生应答结束至教师反应), 指出后者被研究得更少而可能更重要。Tobin (1987) 的综述整理了此后十余年的结果, 认为三至五秒是较为稳健的区间, 并指出效应在高认知层次任务上更明显。Ingram 与 Elliott (2016) 则从课堂互动分析的角度指出, 等待时间的效应受到教师话轮转换规范的强烈调节——学生是否发言, 取决于他们如何理解此刻的发言权归属。

Ingram 与 Elliott 的结论距本研究的关切最近: 它已经把”归属”引入了讨论。但其分析对象是**话轮的分配规则**, 仍属于有主的应答窗内部的分配问题; 等待时间的整个框架预设了教师提出了问题并在等待答案, 即启动权始终在教师一侧。本研究要检验的恰恰是启动权不在任何一侧的情形。

2.2 孵化效应

Wallas (1926) 的四阶段说将孵化列为创造过程的必经环节。Smith 与 Blankenship (1991) 以固着一遗忘假说解释孵化: 间断期使误导性线索的可及性下降, 从而解除固着。Sio 与 Ormerod (2009) 对 117 项研究的元分析确认了孵化效应的存在(总体效应量偏小但稳定), 并报告了两个调节变量: 间断期任务的认知负荷较低时效应更强; 发散性问题上的效应强于收敛性问题。Dodds、Ward 与 Smith (2012) 的综述则强调实验结果的高度异质性。

孵化研究与本研究的核心分歧在机制主张上, 且方向相反。孵化的主流解释——无意识加工、选择性遗忘、线索重构——共同预设间断期**内有加工在进行**。本研究所界定的无主时段, 其功能主张恰恰不依赖任何加工的发生, 而依赖启动权的悬置。二者可通过在间断期填入占用工作记忆的无关任务加以区分(详见 6.3)。

2.3 间隔效应与反馈时机

自 Ebbinghaus (1885) 以来, 分散练习优于集中练习是学习科学中重复性最好的发现之一。Cepeda 等 (2006) 对 254 项研究的元分析确立了间隔与保持的关系, 并指出最优间隔随保持区间延长而延长。Rohrer 与 Taylor (2006) 在数学学习上复现了这一结论。

反馈时机的研究则呈现长期未决的分歧。Kulik 与 Kulik (1988) 的元分析报告即时反馈在多数应用场景中优于延迟反馈; Butler、Karpicke 与 Roediger (2007) 与 Butler 与 Roediger (2008) 则在多项选择测验的保持测验上报告延迟反馈的优势。Shute (2008) 与 Hattie 与 Timperley (2007) 的综述均将此列为尚未整合的分歧, 通行的处理是诉诸任务类型或学习者水平的调节作用。

这一分歧对本研究具有特殊意义: 若作答潜伏期与学习结果的关系是倒 U 型的, 则两侧的实验条件可能分别落在极值点的两侧, 从而两组结果可以同时为真。本研究在 6.2 节展开这一整合解释。

2.4 提示与帮助寻求研究

智能辅导系统中的提示机制为本研究提供了直接可用的数据结构。Aleven 等（2003）系统梳理了交互式学习环境中的帮助寻求行为，指出存在”帮助滥用”与”帮助回避”两类失调。Baker 等（2004）刻画了”系统博弈”行为——学习者通过快速连续请求提示以获得答案。Roll 等（2011）尝试通过元认知辅导改善帮助寻求。Beck（2005）提出以作答时间建模学习者的脱离状态。

已有研究一致地将提示视为一种**支持**，讨论集中于如何使学习者恰当地使用它。本研究引入的视角不同：提示同时是一种**接管**，它把下一动作的启动权从学习者一侧移至系统一侧。这一视角使得”提示使用”可以被读作外部起搏强度的代理量。

2.5 学习者建模中的作答时间

Corbett 与 Anderson（1995）的贝叶斯知识追踪与 Piech 等（2015）的深度知识追踪，均以作答正误序列估计掌握概率，作答时间在这些模型中通常不作为核心变量。Feng、Heffernan 与 Koedinger（2009）与 Heffernan 与 Heffernan（2014）对 ASSISTments 平台的介绍表明，该平台在题目层记录了首动作类型与首动作潜伏期，为本研究提供了必要字段。

综上，作答潜伏期在学习分析中长期作为协变量或脱离度指标出现，鲜有研究将其关系形态本身作为待估参数。

2.6 跨领域参照的文献基础

心脏电生理。Mandel 等（1971）与 Narula 等（1972）确立了以超速起搏后的窦房结恢复时间评估起搏功能的临床方法。其要点有三：备份起搏能力仅在外部的起搏撤除之后可测；长期起搏依赖会抑制下级起搏点的自律性；该抑制在起搏工作期间不产生任何症状或指标异常。

燃烧动力学。Semenov（1928）的热爆炸理论与 Frank-Kamenetskii（1955）的工作确立了着火延迟的理论地位。Heywood（1988）系统阐述了压燃机中延迟期与混合过程的关系，并指出延迟期过短导致爆震。Yao、Zheng 与 Liu（2009）对均质压燃的综述表明，该技术路线的核心困难正是延迟期过短而难以控制燃烧相位。工程实践已从”更短”转向”可控”。

微生物学。Monod（1949）确立了生长曲线的分段描述。Baranyi 与 Roberts（1994）的模型将迟滞期长度归因于细胞适应新环境所需完成的工作量，即前后两个环境之差；Swinnen 等（2004）的综述系统评述了迟滞期建模的进展与分歧。这一系列工作的共同含义是：迟滞期既不属于菌株也不属于培养基，而属于两个环境之间的关系。

2.7 述评

上述五类教育研究文献与三类跨领域文献，共同指向一个未被检验的预设：

没有输出的那一段时间，是待压缩的成本。

该预设在教育研究中不以主张的形式出现，而以指标选取的形式存在。它包含三层：其一，时间被分为有产出与无产出两类且此二分穷尽；其二，无产出的一类被赋予单一负号；其三，无产出时间的价值由其两端的事件决定，其自身不携带信息。

第三层是本研究的靶点。它直接导致记录系统不为该段时间设置字段；而字段的缺失使压缩它的代价在原则上不可被观察到，从而使压缩行为永远不会遭遇反例。本研究称此结构为**压缩的自证**。

值得注意的是，三个跨领域参照中已有两个（燃烧与微生物）在工程实践上撤回了单调压缩的取向，转向“控制在窗口内”。教育技术尚未走到这一步。

2.8 教育时间研究的谱系与本研究的定位

将上述文献置于一个谱系中，可以看清本研究所处的位置。教育研究对“时间”的处理大致可分为四层。

第一层，时间作为资源。学习时间模型（Carroll 传统）与学业投入时间研究把时间视为可分配的稀缺资源，核心问题是“分配多少”。在这一层，无产出时间必然为损耗，因为它被分配却未产生学习的部分。

第二层，时间作为间隔。间隔效应与孵化效应把时间视为两个事件之间的距离，核心问题是“隔多久”。这一层第一次承认无产出时间可能有正面价值，但其价值仍由两端的事件定义——间隔之所以有用，是因为它改变了第二次学习的条件。

第三层，时间作为节奏。等待时间研究与课堂互动分析把时间视为交互序列中的节拍，核心问题是“谁在什么时候说话”。这一层引入了参与者结构，因而最接近本研究。Ingram 与 Elliott（2016）已经指出学生是否发言取决于其对发言权归属的理解，这实际上把归属问题提了出来。

第四层，时间作为归属状态。这是本研究主张的层次：一段时间的性质，由这一段里谁持有下一动作的启动权决定，而不由其长度、位置或两端的事件决定。

前三层的共同点是：它们都可以在不追问归属的情况下完整表述。第三层虽已触及参与者结构，但其分析对象始终是有主的**应答窗内部的分配问题**——教师提出了问题，问题在等待回答，分歧只在于谁来回答、何时回答。本研究要处理的是应答窗尚未建立、或已被明确解除的情形。

这一定位同时说明了本研究可能的失败方式：如果第四层不能给出前三层给不出的可观测预测，那么它就不是新的一层，而只是第三层的一种措辞。6.3 节给出的两因子设计正是为回答这一质疑而设。

3 理论框架与研究假设

3.1 无主时段的概念界定

定义 无主时段（unclaimed interval）指同时满足以下三个条件的一段时间：（a）上一程序已交出控制；（b）下一程序尚未接手；（c）任何一方均不持有对下一动作的启动权。

三个条件缺一不可。条件（a）排除了同一程序内部的迟缓；条件（b）排除了动作已确定而尚未执行的准备状态；条件（c）是本定义的核心，也最易被误读。它并非指“无人在场”，而是指**没有任何一**

方对该段时间的用途拥有主张权。教师在旁等待学生想出答案，不满足条件（c）——启动权在教师一侧，学习者欠其一个应答；学习者内心持续自我催促，同样不满足——启动权在其自我监控成分一侧。

判定问话 为使该定义可在现场使用，本研究给出一条不依赖当事人自述的判定问话：

在这一段时间里，若什么都不发生，谁会被认为失职？

若存在这样的一方，则该段不是无主时段；若不存在，则是。该问话将判定从心理状态转移到责任归属结构上，因而具备观察者间一致性的可能。

三种形态 无主时段在教学情境中呈现三种形态：（I）**交接空档**，上一动作已完成并被确认，下一动作尚未被任何一方指定；（II）**受阻后的空档**，上一程序失败并交出控制，下一程序未接手；（III）**撤除后的空档**，外部供给刚被撤除，学习者尚未自行启动。三者中仅形态III可由外部安排产生，形态I与II只能被不打破，不能被制造。

3.2 与近邻概念的辨析

表 1 给出无主时段与七个近邻概念的分离条件。每一条均给出可判定的观察场合，而非仅作语义区分。

表 1 无主时段与近邻概念的分离条件

近邻概念	该概念的核心主张	分离条件（可判定场合）
休息	为恢复而中断	休息有持有者，事后可结算为”已恢复”；无主时段不可结算
犹豫	在选项之间比较	犹豫表明下一程序已接手，违反条件（b）
迟缓	同一程序执行得慢	控制未被交出，违反条件（a）
中断	外因打断	外因即新的持有者，违反条件（c）
等待时间	教师提问后延长停顿	启动权在教师一侧；延长等待并加入引导性提示后，两理论预测分离
孵化	间断期内的无意识加工	孵化要求间断期内有加工发生；填入占用工作记忆的任务后，两理论预测分离
组织冗余	未被指派用途的资源	冗余有持有者，可回答”这是谁的冗余”；无主时段按定义无法回答该问题

3.3 概念的可操作化：三种形态、清点手册与观察者一致性

一个概念若要进入经验研究，需要一份使不同观察者得出相同判断的清点规则。本节给出该规则。

三种形态的正例、反例与误判阻挡。

形态 I（交接空档）。上一动作已完成并被系统或教师确认，下一动作尚未被任何一方指定。 正例：一道题作答完毕，系统未自动推进，学习者亦未点击继续。 反例：系统显示”3 秒后进入下一题”——该 3 秒的持有者是系统。 误判阻挡：不得将系统加载时间计入；加载期间启动权在系统一侧。

形态 II（受阻后的空档）。上一程序失败并交出控制，下一程序尚未接手。 正例：学习者尝试一次未成功，随后既未再尝试、未请求提示，亦未转向他题。 反例：学习者持续在心中演算——同一程序仍在运行，属迟缓。 误判阻挡：**形态 II 与迟缓在外部不可分辨。**这是本清点手册最主要的缺口，也是 7.1 节所述代理变量不完全的具体形态。在无法区分时，应记为”不可判定”，而不应默认归入任一类。

形态 III（撤除后的空档）。外部供给刚被撤除，学习者尚未自行启动。 正例：教师宣布本段不讲授、不提问、不布置产出，随后的时间。 反例：教师宣布”自己想想，五分钟后我要提问”——该五分钟有持有者。 误判阻挡：只要该段末尾挂有任何形式的交代，整段均不计为无主时段，无论其间学习者的行为如何。

判定的两步程序。观察者按以下顺序执行，不得跳步：

第一步，确认上一程序是否已交出控制（形态 I 为完成，形态 II 为失败，形态 III 为撤除）。若否，记为迟缓。 第二步，回答判定问话——若这一段什么都不发生，谁会被认为失职。存在这样一方则记为有主段；不存在则记为无主时段；无法确定则记为不可判定。

观察者间一致性的预期。该程序把判断从心理状态转移到责任归属结构，理论上具备较高的一致性。但本研究未对其进行实测：本研究使用的是既有平台数据，不存在人工编码环节。附录 C 的实验方案中包含双编码者一致性的检验，且规定在一致性未达标之前，其余检验一律不执行。

三种形态的可安排性差异。形态 III 可由外部安排产生，形态 I 与形态 II 只能被不打断，不能被制造。这一差异具有实践后果：一个体系保护无主时段的能力，其上限等于它安排形态 III 的能力；而形态 I 与 II 的存续，完全取决于在场者的克制。前者可写入课程表，后者不能。8.2 节的两条建议正落在这一分界的两侧。

3.4 三条命题

命题 P1（抑制命题） 外部按时供给的持续强度越高，供给撤除之后学习者自主启动所需的潜伏期越长。形式化为

$$\tau = \tau_0 (1 + kT), k > 0$$

其中 T 为外部起搏强度， τ 为撤除后的自启动潜伏期。该命题移植自超速抑制的临床形式。

命题 P2（非零下界命题） 存在一个大于零的最小无主时段 δ^* ，使学习结果关于无主时段长度的函数在 δ^* 处取极大；低于 δ^* 不构成量的改善，而是切换至另一种过程类型。该命题移植自燃烧学中延迟期与爆震的关系。

命题 P3（无主性守恒命题） 无主时段一旦被赋予持有者——被命名、被计入记录、被要求事后交代、被设为教学环节——即转化为具有产出要求的时间，其功能随之消失。该命题移植自迟滞期的归属性质，并与 Campbell（1979）、Espeland 与 Sauder（2007）关于指标反应性的论述相容。

P3 有一个不利于本研究自身的推论：任何以提高无主时段质量为目的的干预，只要包含记录或评价，即消灭其对象。这一推论在第 7 部分作为结构性局限重申。

3.5 研究假设

三条命题在智能辅导系统的数据结构上转化为四条可检验假设。转化所依据的映射见表 2。

表 2 跨领域概念到数据字段的映射

理论概念	电生理对应物	数据字段操作化
外部起搏	起搏脉冲	曝光窗内首动作为”请求提示”的题目比例 T
自主启动	自主搏动	首动作为”直接作答”的题目
撤除后潜伏期 τ	逸搏间期	检验窗内自主作答题目的首动作潜伏期中位数
无主时段（代理）	——	首动作潜伏期

H1（源自 P1） 控制学习者能力后，曝光窗内的外部起搏强度 T 与检验窗内的自启动潜伏期 τ 正相关。判定：T 的回归系数 $\beta > 0$ 且 $p < 0.01$ 。

H2（源自 P2） 在题目层，学习者内标准化的对数作答潜伏期 z 与正确率呈倒 U 型，且极值点位于 z 分布的第 5 与第 95 百分位之间。判定：二次项系数为负且 $p < 0.01$ ，同时极值点落在上述区间内。

H3（撤销规则） 若正确率关于 z 呈单调递减，即线性项显著为负而二次项不显著、或极值点落在第 5 百分位以下，则 P2 被证伪，本研究的核心主张撤销。

H4（无主性的行为判据） 在同等长度的潜伏期中，启动权归属不同则结果不同。判定：自主作答组的当题正确率显著高于请求提示组，且其在同一技能下一次作答的正确率增益亦显著更高。

需要强调，H3 是一条**自毁开关**：本研究在观察任何结果之前即写定了使自身核心主张失效的条件。按 Nosek 等（2018）的论述，此类条件是预注册区别于事后合理化的关键要件。

4 研究方法

4.1 研究设计

本研究采用**预注册的观察性设计**。选择观察设计而非实验设计的理由有二。其一，本研究关注的关系形态（倒 U）需要在潜伏期的宽分布上估计，实验室条件下难以获得足够的分布覆盖。其二，本研究

要检验的核心结构涉及数月的起搏暴露，实验操纵在伦理与可行性上均不成立。观察设计的代价是无法支持因果推断，本研究在结果与讨论部分严格遵守这一限制。

为弥补观察设计在推断上的弱势，本研究采取三项措施：（1）全部假设、阈值与纳入排除标准于观察结果之前写定并封存哈希；（2）在两个独立年份的数据集上以同一份代码复现；（3）预设一条使核心主张失效的撤销规则。

一项判决性实验设计（2×2 被试内交叉）作为本研究的后续检验，其完整方案见附录 C。

4.2 数据来源与样本

数据取自 ASSISTments 平台的两个公开数据集：2009—2010 学年的 skill builder 数据集与 2012—2013 学年数据集。该平台为中学数学的智能辅导系统，在题目层记录首动作类型、首动作潜伏期、尝试次数、提示次数与正误（Feng 等，2009；Heffernan & Heffernan，2014）。

纳入排除标准（预注册）：仅保留主问题（original = 1），剔除脚手架子题；剔除首动作潜伏期非正值或超过 600 000 毫秒（10 分钟）的记录；要求学习者在曝光窗与检验窗各具备至少 20 条有效记录。

表 3 样本筛选过程

步骤	2009—2010	2012—2013
原始记录	401 756	5 819 737
保留主问题	328 291	5 819 737
潜伏期有效	326 384	5 819 312
双窗各 ≥ 20 条	286 228	5 468 190
学习者数	1 440	24 860

两个数据集来自同一平台的不同学年与不同学习者群体。本研究不主张二者构成完全独立的两次取样——平台界面、提示机制与计时口径相同——仅主张其排除了单一学年的偶然性。这一限制在第 7 部分重申。

4.3 变量操作化

窗口划分。每名学习者的记录按平台记录顺序升序排列，前 50% 划为曝光窗，后 50% 划为检验窗。该比例在预注册中写定，分析过程中未作调整。窗口划分的作用是使”起搏暴露”与”撤除后表现”在时间上分离。

外部起搏强度 T。曝光窗内首动作为”请求提示”的题目所占比例，取值于 [0, 1]。

自启动潜伏期 τ 。检验窗内首动作为”直接作答”的题目上，首动作潜伏期的中位数（毫秒）。取中位数而非均值，以降低极端值影响。

学习者内标准化潜伏期 z 。对首动作潜伏期取自然对数后，在每名学习者内部作标准化。该步骤为必需：学习者之间的基线速度差异极大，不作标准化则倒 U 型关系会被个体间差异掩盖。其代价是本研究只能讨论“相对于该学习者自身的快慢”，无法给出绝对秒数的建议。

能力。曝光窗内的正确率。

题目难度。技能层的平均正确率。

4.4 分析策略

H1 采用最小二乘回归：

$$\log(\tau) = \beta_0 + \beta_1 T + \beta_2 \cdot \text{能力} + \beta_3 \cdot \log(\text{曝光窗题数}) + \varepsilon$$

H2 与 H3 采用二次逻辑回归：

$$\text{logit } P(\text{正确}) = \gamma_0 + \gamma_1 z + \gamma_2 z^2 + \gamma_3 \cdot \text{尝试次数} + \gamma_4 \cdot \text{技能难度}$$

极值点由 $z^* = -\gamma_1 / (2\gamma_2)$ 求得。纳入技能难度的理由是：难题同时具有较低正确率与较长潜伏期，若不加控制，倒 U 的右支可被难度完全解释。

H4 采用两比例 z 检验，比较长潜伏段内自主作答组与请求提示组的正确率。

极端值处理：标准化后剔除 $|z| \geq 6$ 的记录。该规则在预注册中写定。

4.5 预注册与可复现性

预注册文本在任何结果被观察之前写定，其 SHA-256 摘要为：

```
66229cc931778e18314c06b24df1bab675fd488174cdb13e9de92b900bd19dc5
```

预注册明确规定：四条假设全部报告，不作选择性报告；不因结果不佳更换窗口比例、剔除阈值或统计量；事后产生的分析一律标注为探索性。

复算路径见附录 D。分析脚本不含随机数，不含依结果分支的逻辑，输出可逐位比对。2012—2013 数据集的原始文件约 30 亿字节，本研究先按预注册的字段清单作流式抽取，抽取过程不含任何依结果的筛选。

4.6 伦理

两个数据集均为公开发布的去标识化数据，不含可识别个人的信息，使用符合原发布方的条款。本研究未接触任何原始个人信息，无需伦理审查。附录 C 的判决性实验方案包含完整的伦理条款，其实施须另行取得审查批准。

4.7 统计功效与敏感性

功效。H2 的检验在题目层进行，两个数据集的有效观测分别为 264 612 与 2 452 773 条。在此规模下，二次项的检出功效远高于常规阈值，因而显著性本身不构成有力证据——关键在于效应的形态与

跨数据集的一致性，而非 p 值。本研究因此在报告中侧重于系数大小、极值点位置及其在两个独立年份上的复现程度。

H1 的检验在学习者层进行，有效样本为 1 430 与 24 857。在 $\alpha = 0.01$ 、 $1-\beta = 0.80$ 下，前者可检出标准化回归系数约 0.09 以上的效应，后者约 0.02 以上。H1 在较小样本上的不显著 ($p = 0.247$) 与在较大样本上的显著反向 ($p = 8.1 \times 10^{-10}$) 之间不构成矛盾：两者的点估计同号 (-0.210 与 -0.340)，差别在于精度。

敏感性。本研究就三项预处理选择作了敏感性检查。（1）极端值剔除阈值由 $|z| \geq 6$ 收紧至 $|z| \geq 4$ ，二次项的符号、量级与显著性均不变。（2）技能难度代理纳入与否，二次项系数变化小于 0.01。（3）窗口比例为预注册所固定，未作变动；本研究不报告其他比例下的结果，以免构成事后选择。

多重比较。本研究共检验四条预注册假设，全部报告，不作选择性呈现。由于四条假设检验的是不同命题而非同一命题的多个操作化，本研究未施加族系错误率校正；即便按 Bonferroni 校正至 $\alpha = 0.0025$ ，H2 与 H1 在大样本上的结论亦不改变。

5 研究结果

5.1 描述统计

两个数据集的入选样本规模见表 3。2009—2010 数据集中，学习者的外部起搏强度 T 的中位数为 0.009，四分位区间为 [0, 0.050]；2012—2013 数据集中 T 的中位数为 0.000，四分位区间为 [0, 0.031]。两数据集中绝大多数学习者极少请求提示，T 呈强右偏分布，这一点在解释 H1 时须予考虑。

检验窗内自主作答的首动作潜伏期中位数，2009—2010 为 23 千余毫秒，2012—2013 为 25 千余毫秒，两者接近。

5.2 H1：未获支持，且方向相反

表 4 H1 回归结果（因变量：log 自启动潜伏期）

预测变量	2009—2010 (n = 1 430)	2012—2013 (n = 24 857)
外部起搏强度 T	-0.210 (SE = 0.181, p = 0.247)	-0.340 (SE = 0.055, p = 8.1×10^{-10})
能力	+0.110 (p = 0.380)	-0.037 (p = 0.243)
log 曝光窗题数	+0.001 (p = 0.970)	-0.071 (p = 3.1×10^{-44})

H1 预测 T 的系数为正。实测在两个数据集上均为负，且在较大样本上高度显著。**H1 未获支持，且方向与预测相反。**

按预注册纪律，本研究不修改假设、不更换统计量、不追加协变量以挽救该结果。命题 P1 的原始表述在本数据上被推翻。改判后的表述见 6.1 节。

5.3 H2：在两个数据集上一致获得支持

表 5 H2 二次逻辑回归结果（因变量：当题正确）

参数	2009—2010（n = 264 612）	2012—2013（n = 2 452 773）
z（线性项）	-0.061（p = 7.0×10 ⁻²⁹ ）	-0.010（p = 1.1×10 ⁻⁶ ）
z ² （二次项）	-0.152（p < 10 ⁻³⁰⁰ ）	-0.135（p < 10 ⁻³⁰⁰ ）
极值点 z*	-0.200	-0.038
z 的第 5／95 百分位	-1.56／+1.70	-1.48／+1.75

二次项在两个数据集上均显著为负，极值点均位于学习者自身分布的内部，远高于第 5 百分位。**H2 获得支持。**

表 6 给出按 z 十分位的正确率（探索性描述）：

表 6 按标准化潜伏期十分位的正确率（2009—2010）

十分位 （最快 →最 慢）	1	2	3	4	5	6	7	8	9	10
正确率	.699	.748	.737	.725	.707	.683	.666	.636	.604	.565

峰值出现在第二个十分位而非第一个。作答最快的十分之一，其正确率较第二档低 4.9 个百分点。这一形态即 $\delta^* > 0$ 的经验表现。

同时须注意，最慢的十分位正确率为 .565，为全部十档最低，较最快档低 13.4 个百分点。**该数据同样否定“越慢越好”的读法。**本研究主张的是关系的非单调性，而非慢优于快。

5.4 H3：撤销规则未被触发

单调递减关系不成立：二次项极显著且极值点位于分布内部。撤销规则未触发，P2 保留。

5.5 H4：因定义性绑定而不可解释

预注册的 H4 比较长潜伏段中自主作答组与请求提示组的当题正确率。实测请求提示组的正确率为 **0.0000**（2009—2010，n = 1 983；2012—2013，n = 32 285），非接近于零而是恰为零。

原因在于本研究未核对字段语义：在该平台中，首动作为”请求提示”与该题正误记为 0 之间存在**定义性绑定**——请求提示后该题不再计为首次作答正确。因此该检验在设计上不可能得到其他结果。

H4 不作”未获支持”处理，而作”不可解释”处理并予作废。该缺陷由本研究自身的操作化失误造成，平台文档中已有说明，本研究在预注册阶段未予核对。此处如实报告，因为一份预注册中若有一条假设在设计上即不可能成立，其影响不限于该条假设本身。

5.6 稳健性检验

协变量的影响。在 H2 模型中纳入与不纳入技能难度代理，二次项系数变化甚微（2009—2010：-0.152；2012—2013：-0.135），表明倒 U 型关系并非难度所致。

极端值。将 $|z|$ 的剔除阈值由 6 收紧至 4（探索性），二次项系数的符号与显著性不变。

跨数据集一致性。两个数据集在样本规模上相差约 19 倍，二次项系数相差 0.017，极值点位置均在分布内部。

5.7 探索性分析

以下分析均在观察 H1 结果之后设计，**全部标注为探索性**，其证据地位低于预注册结果。

(1) 链条分析。若外部起搏并非延长潜伏期而是压缩无主时段本身，则应观察到：起搏强度越高，潜伏期越短；潜伏期越长，成绩越好。

表 7 链条分析（探索性）

路径	2009—2010	2012—2013
① $T \rightarrow \log(\text{潜伏期})$	$\beta = -0.210$ ($p = 0.246$)	$\beta = -0.346$ ($p = 4.8 \times 10^{-10}$)
② $\log(\text{潜伏期}) \rightarrow \text{检验窗正确率 (控能力)}$	$\beta = +0.008$ ($p = 0.156$)	$\beta = +0.003$ ($p = 3.5 \times 10^{-3}$)
③ $T \rightarrow \text{检验窗正确率 (控能力)}$	$\beta = -0.171$ ($p = 7.4 \times 10^{-6}$)	$\beta = -0.194$ ($p = 3.1 \times 10^{-83}$)

三条路径的方向与改判后的表述一致，但须同时报告三点不利证据。

其一，①与②仅在大样本上显著，效应量甚小。②的系数意味着潜伏期提高一个自然对数单位，正确率仅提高约 0.3 个百分点；该量级不足以支持任何实践建议。

其二，①的四分位描述**非单调**。2012—2013 数据中，T 由最低至最高四分位所对应的潜伏期中位数依次为 23 079、24 584、27 696、26 328 毫秒。回归系数为负而分位描述非单调，说明该关系的形态尚不清楚。

其三，③几乎必然受能力混淆，观察数据无法将其分离。

(2) H4 的修正版。在不受定义性绑定影响的结局上重做比较：长潜伏段中自主作答组与请求提示组在同一技能下一次作答的正确率。

按能力十分位分层（2009—2010），十层中九层为正，加权平均差为 +0.263；唯一反向的一层为能力最高层（-0.015）。学习者内配对（同一学习者兼有两类记录且各不少于 3 条， $n = 215$ ）的均差为 +0.177 ($t = 7.77$, $p = 7.6 \times 10^{-15}$)，其中 67.4% 的学习者差值为正。

学习者内配对排除了个体层面的能力混淆，是本研究实证部分方向性最明确的一项结果。但其仍为观察性证据：学习者选择停留或求助，该选择本身可能由题目难度、疲劳或时间压力决定。本研究不作因果解释。

5.9 δ^* 的边界条件：按能力、尝试次数与难度的分组分析（探索性）

命题 P2 只主张 δ^* 存在且大于零，未对其取值作任何主张。6.4 节第（2）条同时指出，燃烧学从不给出跨工况的最优延迟值，而是给出每组工况下估计它的方法。据此，本研究在两个数据集上分别按三个变量分组重估极值点。以下分析在观察 H2 结果之后设计，全部标注为探索性。

表 9 按能力四分位的二次项与极值点

能力四分位	2009—2010 n	z^2 系数	极值点 z^*	2012—2013 n	z^2 系数	极值点 z^*
Q1（最低）	66 153	-0.204	+0.371	613 194	-0.172	+0.368
Q2	66 153	-0.170	-0.468	613 193	-0.132	-0.239
Q3	66 153	-0.110	-1.173	613 193	-0.093	-0.968
Q4（最高）	66 153	-0.058	-2.566	613 193	-0.066	-1.936

（二次项在全部八个分组中 $p < 10^{-7}$ ）

该模式在两个数据集上高度一致：**极值点随能力上升而单调左移**。对能力最低的四分之一学习者，最优作答潜伏期高于其自身均值（ $z^* \approx +0.37$ ）；对能力最高的四分之一，极值点落在 -2.6 与 -1.9，已在或低于其分布的第 5 百分位，意味着在其可观测范围内，关系实际上表现为**单调递减**——对这部分学习者，更快确实更好。

这一结果对本研究是双向的。一方面它给出了 δ^* 的第一个系统性调节变量，使 P2 从“存在一个正的下界”细化为“下界随能力上升而下降”；另一方面它**限制了 P2 的适用人群**：对高能力学习者，倒 U 的左支落在观测范围之外，本研究的核心主张在其身上无经验内容。

表 10 按尝试次数与技能难度带的极值点

分组	2009—2010 n	z^2 系数 (p)	z^*	2012—2013 n	z^2 系数 (p)	z^*
首次尝试即作答	217 843	-0.151 ($<10^{-300}$)	-0.214	1 996 289	-0.135 ($<10^{-300}$)	-0.036
尝试 ≥ 2 次	46 442	-0.044 (0.309)	+2.427	455 793	+0.003 (0.933)	—
难题（技能正确率 $< .5$ ）	10 370	-0.425 ($<10^{-70}$)	+1.091	58 682	-0.275 ($<10^{-233}$)	+1.216
中等	156 277	-0.176 ($<10^{-290}$)	-0.035	1 036 448	-0.159 ($<10^{-300}$)	+0.257
易题（技能正确率 $> .7$ ）	97 960	-0.082 ($<10^{-26}$)	-1.793	1 357 630	-0.115 ($<10^{-300}$)	-0.694

两项结果同样在两个数据集上复现。

其一，倒 U 是首次作答的性质，不是受阻过程的性质。在尝试次数不少于两次的记录上，二次项在两个数据集中均不显著（ $p = 0.309$ 与 0.933 ）。这构成对 P2 的一项重要边界：本研究所刻画的关系存在于学习者尚未进入反复试错的那一次动作上；一旦进入试错，潜伏期与正确率之间不再呈倒 U。

其二， δ^* 随题目难度上升而右移。难题上的极值点为 $+1.09$ 与 $+1.22$ ，易题上为 -1.79 与 -0.69 。方向在两个数据集上一致，量级相近。

综合表 9 与表 10，可以得到一个尚未被预注册、但在两个独立年份上一致复现的经验规律：

δ^* 随学习者能力上升而下降，随题目难度上升而上升。

这一规律与 6.2 节的整合解释直接相关：若最优潜伏期同时受能力与难度调节，则反馈时机研究中互相冲突的结果，可能部分地源于各研究样本在能力与难度上的分布差异。这为该分歧提供了一个比”任务类型”或”学习者水平”更具体、且可预先计算的解释路径。

须强调，本节全部结果为探索性，未经预注册，且同样受观察设计的因果限制。其价值在于提出可被后续研究预注册检验的具体假设，而非确立结论。

5.10 结果小结

获得支持：作答潜伏期与正确率呈倒 U 型，极值点不在最快端（两个数据集，共 575 万条记录）。

被推翻：命题 P1 的原始表述。

未获检验：命题 P3。本研究测量的是潜伏期长度，而非无主性本身；按 P3，一旦为无主性设置字段，被记录者即已改变性质。这一限制并非取样不足所致，而是该命题对本研究自身实证部分的直接判决。

作废：H4。

6 讨论

6.1 主要发现与命题改判

本研究最重要的一项发现是一个不利结果：命题 P1 被本研究自身的预注册检验推翻，且方向相反。

该结果的含义并非”外部起搏无害”。结合链条分析，一个更符合数据的表述是：

命题 P1'（改判） 外部起搏并不延长撤除后的自启动潜伏期，而是直接压缩无主时段本身。高度依赖外部供给的学习者，其行为收敛为二值模式——立即作答或立即求助——其间的那一段消失。

P1' 与 P1 的差别不是程度而是位置：P1 假定被损害的是**启动的速度**，P1' 主张被损害的是**启动之前那段时间不属于任何一方的时间的存在**。就本研究关心的问题而言，P1' 是更难被满足的主张，因为它要求潜伏期分布的形态发生改变，而不仅是中心位置移动。

P1' 的证伪条件（新提出，未经预注册，标注为待检验）：测量作答潜伏期分布的离散度与形态而非中位数。P1' 预测高起搏依赖组的分布更集中于两端（呈双峰），低依赖组更接近单峰。若高依赖组同样呈单峰且方差不更小，P1' 被证伪。该检验可在本研究所用的两个数据集上完成；本研究未予执行，因其系在观察结果之后提出，按预注册纪律不得作为预注册假设报告。

6.2 对反馈时机之争的理论意涵

倒 U 型关系为反馈时机的长期分歧提供了一个可检验的整合解释。

如 2.3 节所述，即时反馈与延迟反馈各有元分析支持，通行的调节变量解释（任务类型、学习者水平）无法预先给出任何可计算的量。若学习结果关于作答潜伏期呈倒 U，则两组结果可以同时为真：**支持即时反馈的实验条件多落在极值点左侧，支持延迟反馈的条件多落在右侧或跨越极值点**。分歧之所以持续，是因为双方均在检验”更早或更晚孰优”，而无人估计”极值点位于何处”。

这一解释对教育研究提出了一项具体的、其自身框架内未曾提出的测量任务：**将极值点作为待估参数，而非将时机作为二分处理变量**。该任务的形式约束来自燃烧学——工程领域从不给出跨工况的最优延迟数值，而是给出每组工况下估计它的方法。本研究采取同一立场：极值点的位置随平台、题库与群体而变，5.3 节报告的 $z^* = -0.200$ 与 -0.038 不可外推为绝对秒数建议。

6.3 与四类近邻理论的区分：一项判决性实验设计

表 1 给出的分离条件可合并为一项两因子被试内实验，一次判定四类理论。

设计为 **2（时长：短／长）× 2（启动权：有主／无主）**。有主条件下，教师宣布稍后将请人回答，并在时间到时确实点名；无主条件下，教师宣布该段不收取、不提问，随即转身从事其他工作，且事后不再提及。两条件的时长与题目完全相同，仅归属不同。

表 8 四类理论对两因子设计的对照预测

	等待时间	孵化效应	间隔效应	本研究
时长主效应	显著，长优	显著，长优	显著，长优	不必显著
启动权主效应	无预测	无预测	无预测	显著
交互效应	不显著	不显著	不显著	显著
长·无主 vs 长·有主	相当	相当	相当	前者显著更优
自启动率	无预测	无预测	无预测	前者显著更高

第三行与第四行上四者给出相反的可观测预测，因而一次实验可以判完。若交互效应不显著，本研究的核心区分（有主/无主）失去经验意义，第 3 部分的概念界定应被判为无经验内容。该撤销规则须写入实验的预注册。

完整方案（含逐字指导语、六条执行禁令、拉丁方轮转表、混淆检验、数据表结构与检验力估计）见附录 C。该设计可在一个班级、两周内完成。

6.4 三个参照领域的反向约束

跨领域移植若仅取有利于己的部分，则不构成论证。以下三条为三个参照领域各自对本研究的反向约束，均实质性削弱本研究的主张。

（1）**电生理约束主张强度**。超速抑制是**可逆的**：起搏停止后数分钟至数十分钟内，下级起搏点的自律性多可恢复。据此，本研究不得使用“消除”一类表述，应降格为“抬高启动阈值”，且该抬高在原理上可逆。这一降格改变了问题的性质：若损害可逆，则严重性取决于**撤除机会出现的频率**，而非损害的深度。一个从不撤除外部供给的系统，其风险不在于造成了不可逆损伤，而在于使该损伤永远不获检验的机会。

（2）**燃烧约束可操作性**。最优延迟由燃料性质、压缩比、进气温度、转速与负荷共同决定，燃烧学从未给出跨工况的最优值。相应地，本研究的 P2 只能主张 δ^* 的**存在性**，不能给出取值。任何将本研究读作“应为学习者留出 X 秒”的做法均属误用。

（3）**微生物学约束适用范围**。迟滞期的长度取决于前后环境之差；当差为零时迟滞期趋近于零，而这并非损害。据此，**本研究的主张仅适用于要求迁移与切换的学习**；对于纯熟练化训练（口诀记忆、指法、标准化操作），压缩无主时段不但无代价，且正是目标。

第三条约束切除了本研究在教育实践中相当大的一块适用范围，且本研究**未能给出区分“要求切换的学习”与“纯熟练化训练”的可操作判据**。此为本研究的主要理论缺口之一。

6.5 与制度性测量文献的关系

命题 P3 与 Campbell (1979) 的指标失真论述、Espeland 与 Sauder (2007) 的反应性理论、Scott (1998) 关于国家可读性的论述相容。本研究不与之竞争，而是提出一项增量：被测量行为改变的对象，可以是一段时间的**归属状态**，而限于某一指标的数值。一段时间一旦被记录，它便获得了持有者，而本研究主张其功能恰恰依赖于没有持有者。

这一增量是否成立，取决于一个尚未执行的检验：设计一个记录无主时段但不评价、不反馈、不告知学习者的系统，考察其与学习结果的关系是否因记录而改变（见 7.3）。

6.6 形式移植的合法性问题

本研究从三个自然科学与工程领域引入结构，须回答一个方法论质疑：生理、化学与生物过程的机制与学习过程无关，此类引入是否只是修辞？

本研究的立场是：所移植的不是机制，而是**形式结构与测量策略**，且移植的合法性由以下三条约束保证。

第一，移植的内容必须可独立于原领域表述。本研究从电生理移植的是“备份能力仅在外部供给撤除后可测”，从燃烧移植的是“存在非零下界，低于它是过程类型的切换而非量的改变”，从微生物学移植的是“该时段的长度属于前后两个环境之差而非任一方”。三者均可在不提及心肌细胞、燃料或菌株的情况下完整表述，因而不依赖原领域的机制。

第二，移植必须产生原领域之外的可证伪预测。若移植只产生了对已知现象的重新描述，则属修辞。本研究的三条命题各自产生了教育领域内可检验的预测：P1 产生 H1（已被推翻），P2 产生 H2（已获支持），P3 产生 7.3 节第 6 条的检验设计（尚未执行）。

第三，原领域必须被允许反过来约束本研究。这是最容易被跨域研究省略的一条。6.4 节给出的三条反向约束——主张强度的降格、取值不可外推、适用范围的切除——均实质性削弱了本研究，且第三条切除了教育实践中相当大的一部分。若一项跨域移植只从原领域取得支持而从不接受其限制，则该移植不具备论证地位。

据此，本研究不主张学习过程与心脏起搏、压燃或细菌生长在机制上同构，亦不主张三个参照领域的结论可直接推广至教育。本研究主张的是：三者各自的成熟知识内部均存在与“无产出时间是纯成本”相冲突的结论，且冲突的形式相同；这一形式的重复出现，构成了在教育领域内检验同一形式的理由。

6.7 δ^* 的可变性对整合解释的强化

5.9 节的分组结果强化了 6.2 节提出的整合解释，同时也使其更易被推翻。

强化在于：若 δ^* 随能力上升而下降、随难度上升而上升，则反馈时机研究中互相冲突的结果，可以部分归因于样本在这两个维度上的分布差异。一项以高能力学习者与易题为主的研究，其样本多落在极值点右侧，因而会观察到“更快更好”；一项以低能力学习者与难题为主的研究，其样本多落在左侧，因而会观察到“延迟更好”。该解释给出的不再是一个模糊的调节变量，而是一个可以在研究设计阶段预先计算的位置。

易被推翻在于：该解释作出了一个明确预测——**重新分析既有反馈时机研究，其结果方向应当与样本的能力—难度构成系统相关。**若不相关，本整合解释被证伪。这一再分析不需要新数据，只需要既有研究报告其样本的能力分布与题目难度分布，其成本远低于新的实验。

7 研究局限

7.1 设计与推断上的局限

(1) 观察设计不支持因果推断。本研究全部结果均为相关性证据。5.7 节路径③（起搏强度与检验窗正确率的负相关）几乎必然受能力混淆——频繁请求提示与学业表现较差之间存在过多通路，观察数据无法分离。任何据此得出“应减少提示”之类结论的做法，均属对本研究明确声明不可作因果解释之结果的误用。

(2) **代理变量不完全。**本研究测量的是作答潜伏期，而非无主时段本身。一段较长的潜伏可能是无主时段，也可能是学习者在自我催促下的紧张停留——按 3.1 节的定义，后者不属于无主时段。本研究不具备区分二者的字段，且按命题 P3，此类字段一旦存在，被记录者已非原对象。

(3) **两个数据集非完全独立。**二者来自同一平台，界面逻辑、提示机制与计时口径相同。若在交互逻辑不同的系统（例如允许无限次修改、不记录首动作的系统）上倒 U 型关系消失，则本研究测得的可能是该平台交互设计的性质，而非学习过程的性质。**此为本研究最应受质疑之处。**

(4) **操作化失误。**H4 因字段间的定义性绑定而作废（5.5 节）。该失误可通过核对平台文档避免。其影响不限于该条假设：一份预注册中出现设计上不可能成立的假设，应导致读者对其余部分的可信度作相应折减。

(5) **效应量微小。**5.7 节路径②的系数为 +0.003，即潜伏期提高一个自然对数单位对应正确率提高约 0.3 个百分点。该量级仅在大样本上显著，不足以支持实践建议。倒 U 型关系的证据强度显著高于链条分析。

7.2 三个未能解决的理论问题

(1) **适用范围的判据缺失。**如 6.4 节所述，本研究无法区分“要求切换的学习”与“纯熟练化训练”。微生物学具备相应的量（环境之差），教育研究中尚无对应物。

(2) **发起能力建立阶段的边界缺失。**本研究的建议与发展科学中的回应式照护研究存在正面张力：后者主张成人应及时敏感地回应儿童的发起动作。本研究的处理是划分适用阶段——回应式照护处理发起能力的建立阶段，本研究处理其已建立而被外部供给持续接管的阶段。但本研究未能给出判断“发起程序是否已建立”的可操作判据，该分界目前只能诉诸年龄等粗糙代理。

(3) **结构性两难。**按命题 P3，任何保护无主时段的制度安排均将消灭其对象。这不是执行困难，而是结构性的两难，本研究无解。

7.3 后续研究的六条路径

按成本由低至高排列：

1. **两因子判决性实验**（附录 C）。一个班级、两周。若交互效应不显著，本研究第 3 部分的核心区分失去经验意义。
2. **孵化期填充实验。**在间断期填入占用工作记忆的无关任务，若效果保持，则本研究与孵化理论不可分辨。
3. **P1' 的分布形态检验。**在现有数据上即可完成（6.1 节）。
4. **δ^* 的跨平台一致性检验。**若极值点在三个以上平台上无规律漂移，则 P2 虽成立而无理论内容，退化为平台工程参数。
5. **跨系统复现。**在交互逻辑不同的系统上重做 H2。
6. **P3 的直接检验。**设计记录而不评价、不反馈、不告知的系统。该检验自身存在一个逻辑困难：未告知学习者的记录是否构成“赋予持有者”，取决于持有者是否必须为当事人所知，本研究的定义在这一点上尚不够明确。

7.4 对本研究结论稳健性的总体评估

综合 5.8 节与本部分，本研究各项结论的证据强度可分为四档，读者宜按档次采信。

第一档（较强）：作答潜伏期与正确率呈倒 U 型，极值点不在最快速端。依据为两个学年、575 万条记录上的预注册检验，且在控制难度与尝试次数、变更极端值阈值后保持。主要威胁为两个数据集来自同一平台（7.1 节第 3 条）。

第二档（中等）： δ^* 随能力上升而下降、随难度上升而上升；倒 U 仅存在于首次作答，不存在于反复试错。依据为两个数据集上一致复现的分组分析，但该分析为探索性，未经预注册。

第三档（弱）：外部起搏压缩无主时段本身（P1'）。依据为链条分析，其效应量微小、四分位描述非单调、且仅在大样本上显著。该命题的专门检验（分布形态检验）尚未执行。

第四档（无经验支持）：无主性守恒命题（P3）。本研究未对其作任何直接检验，其地位是一条从三个参照领域的结构推出、并由“记录系统中不存在相应字段”这一事实间接支持的推论。

一项须写明的判断：**本研究的概念贡献（第 3 部分）与其经验贡献（第 5 部分）的强度并不相称。**概念界定覆盖三种形态、七条分离条件与一份清点手册，而经验部分只对其中的一个代理变量作了检验。若后续的两因子实验（附录 C）显示交互效应不显著，第 3 部分的核心区分应当被判为无经验内容，而不应以“概念上有价值”为由保留。

8 结论与教育启示

8.1 主要结论

第一，作答潜伏期与学习结果之间的关系是非单调的，其极值点不位于最快速端。该结论在两个学年、575 万条作答记录上一致成立，并在控制难度与尝试次数后保持。由此，“作答越快越好”这一被广泛内置于教育技术产品中的默认设定不成立。

第二，外部起搏对自主启动的影响，不表现为撤除后潜伏期的延长，而表现为无主时段本身的压缩。原命题被本研究自身的检验推翻，改判后的表述（P1'）尚待专门检验。

第三，无主时段可被界定为一个具有明确边界、可与七个近邻概念分离的独立类别，其判定不依赖当事人自述，而依赖责任归属结构。该界定的经验价值取决于一项尚未执行的两因子实验。

8.2 对教学实践的启示

本研究能够支持的实践含义有限，且必须与 7.1 节的限制一并阅读。

（1）将撤除设为常规而非事故。按 6.4 节第（1）条，风险不在抑制本身，而在于永不撤除。可操作的最小安排是：在课程中定期设置不提供提示、不设反馈时限、不计分的作业段，其唯一目的是使“撤除后学习者多久自行启动”成为可观察的量。代价是该段在全部现行指标上均表现为退步。

(2) **在长潜伏之后不立即介入**。5.7 节的学习者内配对结果给出了方向，但不构成因果证据。据此可提出的默认动作是：当学习者停留时，将原本要做的动作**放慢半拍**——不作反应会被读作回避，介入则构成归属指派。须接受该动作有很高概率不产生可观察效果。

(3) **关于表扬的一项狭窄增量**。不即刻表扬偶发成功这一建议，其解释力大部分已由过度理由效应（Lepper 等，1973；Deci 等，1999）占据。本研究的增量仅在于：表扬不仅引入外部动机，同时构成一次**持有者指派**。区分二者的可判场合是由与该学习者无评价关系的第三方给出同样的表扬——过度理由效应预测同等程度的侵蚀，本研究预测侵蚀较轻。若结果为同等侵蚀，本研究在此点上的增量为零。

8.3 对教育技术设计的启示

(1) **将作答潜伏期从协变量改为设计参数**。现行产品普遍将其作为难度或投入度的代理，本研究表明其与结果的关系形态本身具有设计意涵。

(2) **审查推进机制的默认值**。下一题自动弹出、可见计时、连击奖励等设计，均在压缩形态 I 的交接空档。本研究不主张取消这些设计，主张将“压缩至何种程度”作为需要估计的参数，而非默认取零。

(3) **在系统内实现可撤除性**。智能辅导系统具备实施定期撤除并测量自启动潜伏期的天然条件，这是课堂难以实现而系统可以廉价实现的一项测量。

8.4 对教育评价制度的启示

按命题 P3，任何将无主时段纳入记录与考核的安排都将消灭其对象。因此本研究对制度提出的要求，其形态与通常的教育改革主张相反：通常的主张是**增加**评价维度与路径的多样性，本研究的要求是**允许**账面上存在一栏为空，且该栏不被追查。

这一要求之所以显得激进，不在于其索求甚多，而在于其索求的正是科层运行逻辑的反面。一个仅承认可移交、可复制、可考核之物的组织，其运转速度中并无阅读空白的余地。

一项可行的转译是：不向管理部门提出抽象要求，而以其现行语言争取对应的时间——例如争取每周一节不考试、不提交报告、不布置产出任务的课时。代价是该课时一旦获批，极可能被逐步以监控或汇报的形式收编。

本研究必须写明的自我限制：上述任何一条建议，一旦被写入教师手册、纳入考核或要求汇报执行情况，即按命题 P3 失效。一个只能依靠不被制度化来维持的建议，在教育政策意义上近乎不构成建议。本研究承认这一点，并将其列为第三个未解决的理论问题。

8.5 对教育研究方法的启示

本研究在方法上提供了一个具体的例证，说明预注册在教育研究中的收益并非抽象。

其一，**预注册使一个不利结果具有信息价值**。H1 的失败之所以能够推动理论修订（ $P1 \rightarrow P1'$ ），是因为其预测方向在观察之前即已写定。若无预注册，一个方向相反的显著结果极易被事后重新解释为“起搏使学习者反应更迅速”，从而丧失全部诊断价值。

其二，预注册使自身的设计缺陷暴露。H4 因字段间的定义性绑定而作废（5.5 节）。若无预注册，该比较的结果（一组正确率为零）几乎必然被静默剔除或改写为其他形式的分析，读者无从知晓。作废并公开报告，使读者得以对本研究其余部分的可信度作相应折减——这正是预注册应当提供的信息。

其三，撤销规则的实质意义。H3 规定了一组会使本研究核心主张失效的观测结果。该规则在本次检验中未被触发，但其价值不在于是否触发，而在于它在观察之前就使“何种结果会使本研究失败”成为公开可查的事实。Nosek 等（2018）指出，预注册与事后合理化的分界正在于此。

其四，一项须承认的代价。预注册纪律使 5.9 节的分组分析——本研究方向最清晰、跨数据集一致性最好的一组结果——只能以探索性身份报告，其证据地位低于已被推翻的 H1。这是纪律的成本，本研究接受这一成本，不以任何方式将其升格。

声 明

基金项目 本研究未获任何外部资助。

作者贡献 王德生负责研究构想、理论框架与命题构造；Claude 负责文献整理、数据处理、统计分析与稿件撰写。两位作者共同完成结果解释与修订，并对全文承担责任。

利益冲突 作者声明不存在利益冲突。

数据可得性 本研究使用的两个数据集均为公开数据。预注册文本、其 SHA-256 摘要、全部分析脚本与结果输出随本文一并提供，见附录 A 与附录 D。

生成式人工智能使用声明 本文的文献整理、数据分析与文字撰写由大型语言模型 Claude 参与完成，其贡献已在作者贡献中列明。全部数值结果由可复算的脚本产生，未经人工调整；全部引用文献由作者核对。

致 谢 感谢 ASSISTments 团队将数据公开发布。本研究的问题源自一次关于教学中沉默的长时段讨论，该讨论的记录构成本文第 3 部分概念界定的最初来源。

参考文献

- [1] Aleven V, Stahl E, Schworm S, et al. Help seeking and help design in interactive learning environments[J]. Review of Educational Research, 2003, 73(3): 277-320.
- [2] Baayen R H, Davidson D J, Bates D M. Mixed-effects modeling with crossed random effects for subjects and items[J]. Journal of Memory and Language, 2008, 59(4): 390-412.
- [3] Baker R S, Corbett A T, Koedinger K R, et al. Off-task behavior in the cognitive tutor classroom[C]//Proceedings of ACM CHI 2004. New York: ACM, 2004: 383-390.
- [4] Bandura A. Self-efficacy: Toward a unifying theory of behavioral change[J]. Psychological Review, 1977, 84(2): 191-215.

- [5] Bandura A. Self-Efficacy: The Exercise of Control[M]. New York: W. H. Freeman, 1997.
- [6] Baranyi J, Roberts T A. A dynamic approach to predicting bacterial growth in food[J]. International Journal of Food Microbiology, 1994, 23(3-4): 277-294.
- [7] Beck J E. Engagement tracing: Using response times to model student disengagement[C]//Proceedings of the 12th International Conference on Artificial Intelligence in Education. Amsterdam: IOS Press, 2005: 88-95.
- [8] Berglas S, Jones E E. Drug choice as a self-handicapping strategy in response to noncontingent success[J]. Journal of Personality and Social Psychology, 1978, 36(4): 405-417.
- [9] Bion W R. Learning from Experience[M]. London: Heinemann, 1962.
- [10] Bjork R A. Memory and metamemory considerations in the training of human beings[M]//Metcalfe J, Shimamura A. Metacognition: Knowing about Knowing. Cambridge, MA: MIT Press, 1994: 185-205.
- [11] Bjork E L, Bjork R A. Making things hard on yourself, but in a good way[M]//Gernsbacher M A, et al. Psychology and the Real World. New York: Worth, 2011: 56-64.
- [12] Brown J S, VanLehn K. Repair theory: A generative theory of bugs in procedural skills[J]. Cognitive Science, 1980, 4(4): 379-426.
- [13] Butler A C, Karpicke J D, Roediger H L. The effect of type and timing of feedback on learning from multiple-choice tests[J]. Journal of Experimental Psychology: Applied, 2007, 13(4): 273-281.
- [14] Butler A C, Roediger H L. Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing[J]. Memory & Cognition, 2008, 36(3): 604-616.
- [15] Campbell D T. Assessing the impact of planned social change[J]. Evaluation and Program Planning, 1979, 2(1): 67-90.
- [16] Cepeda N J, Pashler H, Vul E, et al. Distributed practice in verbal recall tasks: A review and quantitative synthesis[J]. Psychological Bulletin, 2006, 132(3): 354-380.
- [17] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. User Modeling and User-Adapted Interaction, 1995, 4(4): 253-278.
- [18] Cyert R M, March J G. A Behavioral Theory of the Firm[M]. Englewood Cliffs, NJ: Prentice-Hall, 1963.
- [19] Deci E L, Koestner R, Ryan R M. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation[J]. Psychological Bulletin, 1999, 125(6): 627-668.

- [20] Dodds R A, Ward T B, Smith S M. A review of experimental research on incubation in problem solving and creativity[M]//Runco M A. Creativity Research Handbook. Cresskill, NJ: Hampton Press, 2012.
- [21] Dweck C S. Self-Theories: Their Role in Motivation, Personality, and Development[M]. Philadelphia: Psychology Press, 2000.
- [22] Ebbinghaus H. Über das Gedächtnis[M]. Leipzig: Duncker & Humblot, 1885.
- [23] Espeland W N, Sauder M. Rankings and reactivity: How public measures recreate social worlds[J]. American Journal of Sociology, 2007, 113(1): 1-40.
- [24] Feng M, Heffernan N, Koedinger K. Addressing the assessment challenge with an online system that tutors as it assesses[J]. User Modeling and User-Adapted Interaction, 2009, 19(3): 243-266.
- [25] Frank-Kamenetskii D A. Diffusion and Heat Exchange in Chemical Kinetics[M]. Princeton, NJ: Princeton University Press, 1955.
- [26] Hattie J, Timperley H. The power of feedback[J]. Review of Educational Research, 2007, 77(1): 81-112.
- [27] Heffernan N T, Heffernan C L. The ASSISTments ecosystem: Building a platform that brings scientists and teachers together[J]. International Journal of Artificial Intelligence in Education, 2014, 24(4): 470-497.
- [28] Heywood J B. Internal Combustion Engine Fundamentals[M]. New York: McGraw-Hill, 1988.
- [29] Ingram J, Elliott V. A critical analysis of the role of wait time in classroom interactions[J]. Cambridge Journal of Education, 2016, 46(1): 37-53.
- [30] Kulik J A, Kulik C C. Timing of feedback and verbal learning[J]. Review of Educational Research, 1988, 58(1): 79-97.
- [31] Lepper M R, Greene D, Nisbett R E. Undermining children' s intrinsic interest with extrinsic reward[J]. Journal of Personality and Social Psychology, 1973, 28(1): 129-137.
- [32] Mandel W J, Hayakawa H, Danzig R, et al. Evaluation of sino-atrial node function in man by overdrive suppression[J]. Circulation, 1971, 44(1): 59-66.
- [33] Miller R B. Response time in man-computer conversational transactions[C]// Proceedings of the AFIPS Fall Joint Computer Conference. 1968: 267-277.
- [34] Monod J. The growth of bacterial cultures[J]. Annual Review of Microbiology, 1949, 3: 371-394.
- [35] Narula O S, Samet P, Javier R P. Significance of the sinus-node recovery time[J]. Circulation, 1972, 45(1): 140-158.

- [36] Nosek B A, Ebersole C R, DeHaven A C, et al. The preregistration revolution[J]. *Proceedings of the National Academy of Sciences*, 2018, 115(11): 2600-2606.
- [37] Piech C, Bassen J, Huang J, et al. Deep knowledge tracing[C]//*Advances in Neural Information Processing Systems* 28. 2015: 505-513.
- [38] Rohrer D, Taylor K. The effects of overlearning and distributed practice on the retention of mathematics knowledge[J]. *Applied Cognitive Psychology*, 2006, 20(9): 1209-1224.
- [39] Roll I, Aleven V, McLaren B M, et al. Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system[J]. *Learning and Instruction*, 2011, 21(2): 267-280.
- [40] Rowe M B. Wait-time and rewards as instructional variables, their influence on language, logic, and fate control[J]. *Journal of Research in Science Teaching*, 1974, 11(2): 81-94.
- [41] Rowe M B. Wait time: Slowing down may be a way of speeding up[J]. *Journal of Teacher Education*, 1986, 37(1): 43-50.
- [42] Scott J C. *Seeing Like a State*[M]. New Haven, CT: Yale University Press, 1998.
- [43] Seligman M E P. *Helplessness: On Depression, Development, and Death*[M]. San Francisco: Freeman, 1975.
- [44] Semenov N N. Zur Theorie des Verbrennungsprozesses[J]. *Zeitschrift für Physik*, 1928, 48: 571-582.
- [45] Shneiderman B. Response time and display rate in human performance with computers[J]. *ACM Computing Surveys*, 1984, 16(3): 265-285.
- [46] Shute V J. Focus on formative feedback[J]. *Review of Educational Research*, 2008, 78(1): 153-189.
- [47] Sio U N, Ormerod T C. Does incubation enhance problem solving? A meta-analytic review[J]. *Psychological Bulletin*, 2009, 135(1): 94-120.
- [48] Smith S M, Blankenship S E. Incubation and the persistence of fixation in problem solving[J]. *American Journal of Psychology*, 1991, 104(1): 61-87.
- [49] Swinnen I A M, Bernaerts K, Dens E J J, et al. Predictive modelling of the microbial lag phase: A review[J]. *International Journal of Food Microbiology*, 2004, 94(2): 137-159.
- [50] Tobin K. The role of wait time in higher cognitive level learning[J]. *Review of Educational Research*, 1987, 57(1): 69-95.
- [51] VanLehn K. *Mind Bugs: The Origins of Procedural Misconceptions*[M]. Cambridge, MA: MIT Press, 1990.

- [52] Wallas G. The Art of Thought[M]. London: Jonathan Cape, 1926.
- [53] Winnicott D W. The Maturation Processes and the Facilitating Environment[M]. London: Hogarth Press, 1965.
- [54] Yao M, Zheng Z, Liu H. Progress and recent trends in homogeneous charge compression ignition (HCCI) engines[J]. Progress in Energy and Combustion Science, 2009, 35(5): 398-437.

附录 A 预注册要点与摘要

预注册文本于任何结果被观察之前写定，SHA-256

摘要为

66229cc931778e18314c06b24df1bab675fd488174cdb13e9de92b900bd19dc5。要点如下：

1. **数据**：ASSISTments 2009—2010 与 2012—2013，字段清单预先列明。
2. **纳入排除**：仅主问题；潜伏期 (0, 600 000] 毫秒；双窗各不少于 20 条。
3. **窗口**：按记录顺序前 50% 为曝光窗，后 50% 为检验窗。
4. **四条假设与判定阈值**：如 3.5 节所列。
5. **撤销规则**：H3，若正确率关于 z 单调递减，则 P2 证伪，核心主张撤销。
6. **复现规则**：主数据集的每条假设，在第二个数据集上以同一代码、同一阈值重跑；结论不一致时以“不一致”为最终结论。
7. **禁止事项**：不作选择性报告；不因结果更换窗口比例、剔除阈值或统计量；不将事后发现写作预注册假设。

附录 B 变量与字段对照

变量	字段	说明
学习者	<code>user_id</code>	去标识化编号
记录顺序	<code>order_id</code>	用于窗口划分
首动作类型	<code>first_action</code>	0 = 直接作答，1 = 请求提示
首动作潜伏期	<code>ms_first_response</code>	毫秒
正误	<code>correct</code>	0/1； $\triangle$ 与 <code>first_action = 1</code> 存在定义性绑定
尝试次数	<code>attempt_count</code>	上限截断为 10
技能	<code>skill_id</code>	用于计算难度代理与后续同技能作答
主问题标识	<code>original</code>	仅保留取值为 1 者

附录 C 两因子判决性实验方案（摘要）

设计 2（时长：5 秒／60 秒）× 2（启动权：有主／无主），被试内交叉，四轮拉丁方轮转。

操纵 有主条件逐字指导语：“这道题先自己想，等一下我请一位同学说说他的想法。”时间到必须真实点名。无主条件逐字指导语：“这一段不用交，也不会问。”随即转身从事其他工作，不显示计时器，事后不再提及。

六条执行禁令（无主条件期间）：不提问；不提示；不巡视至座位；不长时间注视个别学习者；末尾不收取、不检查、不评论；事后任何场合不提起。第六条最易违反且最致命——事后追问等于回溯地指派持有者。

结局变量 主结局为 48 小时后同技能异题的正确率；次结局为自启动率（自由练习段中主动开始的比例及首次动作延迟）。混淆检验为在任率，由不知假设的观察者在第 15、30、45 秒作瞬时抽样。

分析 题目层混合效应逻辑回归， $\text{正确} \sim \text{时长} \times \text{启动权} + (1|\text{学习者}) + (1|\text{题目})$ ，主检验为交互项。

判定 本研究命中须三条同时成立：交互项 $p < 0.01$ ；长·无主显著优于长·有主（ $p < 0.01$ ）；短·无主与短·有主差异不显著。**撤销规则**：时长主效应与交互项均不显著时，本研究第 3 部分的核心区分判为无经验内容。

检验力 一个班 40 人 × 每格 8 题 = 每格 320 条观测，可检出交互项优势比约 1.5（ $\alpha = 0.01$ ， $1-\beta = 0.80$ ）；建议不少于两个班。

最脆弱处 教室情境中无主条件不可能完全无主——同伴在场本身即构成一种归属。结果为零时，此为首要嫌疑，而非本研究被推翻的证据。

附录 D 复算路径

步骤	内容
1	校验预注册文本的 SHA-256 与正文所载一致
2	自公开镜像下载两份原始数据（63 455 745 字节与 3 009 494 391 字节）
3	运行 <code>an.py <csv> <标签></code> ，输出即为表 4、表 5 与 5.5 节
4	运行 <code>chain.py <csv></code> ，输出即为表 7
5	运行 <code>h4b.py <csv></code> ，输出即为 5.7 节第（2）项

脚本不含随机数，不含依结果分支的逻辑，输出可逐位比对。

附录 E 无主时段清点手册（现场使用）

适用 课堂观察、辅导记录、智能辅导系统的日志编码。

第一步·确认控制是否已交出

观察到的情形	记为
上一动作已完成并被确认	形态 I 候选
上一动作失败且未继续	形态 II 候选
外部供给刚被明确撤除	形态 III 候选
同一动作仍在进行（演算、书写、检索）	迟缓，不计

第二步·判定问话

这一段时间里，若什么都不发生，谁会被认为失职？

回答	记为
某一具体方（教师、系统、同伴、学习者本人的自我要求）	有主段
无任何一方	无主时段
无法确定	不可判定（不得默认归入任一类）

第三步·排除项复核（任一项成立即改判为有主段）

1. 该段末尾挂有交代（提问、收取、检查、评论、事后追问）
2. 计时器可见
3. 观察者或教师在该段内注视特定学习者超过三秒
4. 该段被赋予名称或教学目标（如”自主思考环节”）
5. 该段被要求记录、汇报或计入任何评价

第四步·记录字段

字段	取值
形态	I / II / III / 迟缓 / 不可判定
归属	有主 / 无主 / 不可判定
时长	秒
终止方式	学习者自行启动 / 外部介入 / 时间到
排除项	触发的编号（可多选）

两条使用禁令

1. 本手册不得用于评价教师或学习者。按命题 P3，一旦其记录进入评价，被记录的对象即已改变性质，记录随之失效。
2. 不得在事后向学习者询问该段内的心理活动。事后追问构成回溯性的持有者指派，将使该段自始不成立。

已知缺口 形态Ⅱ与迟缓在外部不可分辨（3.3 节）。遇此情形一律记为不可判定。本手册的观察者间一致性尚未实测，其检验方案见附录 C。

附录 F 主要符号与缩写

符号／缩写	含义	首见
T	外部起搏强度，曝光窗内首动作为请求提示的题目比例	3.5 节
τ	撤除后自启动潜伏期，检验窗内自主作答题目首动作潜伏期的中位数	3.4 节
τ_0	无外部起搏时的基线潜伏期	3.4 节
k	抑制系数	3.4 节
z	学习者内标准化的对数首动作潜伏期	4.3 节
z^*	二次逻辑回归的极值点， $z^* = -\gamma_1 / (2\gamma_2)$	4.4 节
δ^*	使学习结果取极大的最小无主时段长度	3.4 节
P1／P1'／P2／P3	抑制命题／改判后的抑制命题／非零下界命题／无主性守恒命题	3.4、6.1 节
H1—H4	四条预注册假设	3.5 节
形态 I ／Ⅱ／Ⅲ	交接空档／受阻后的空档／撤除后的空档	3.3 节