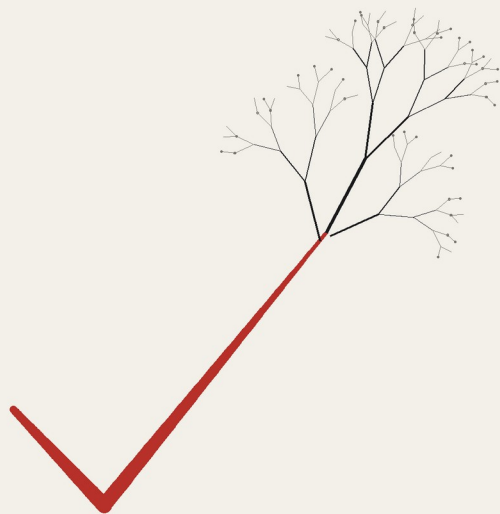




# 答案之后

---

数学学习的智慧

在答案越来越便宜的时代  
重新判断学习有没有真的发生

阳涌 著

Demai International Press

---

答案之后·数学学习的智慧

# 答案之后：数学学习的智慧

著者：阳涌

ISBN 978-1-970820-92-8

德麦国际出版社 Demai International Press

定价：US\$ 10.50

版权所有·侵权必究。本书内容未经著作权人书面许可，不得以任何方式复制、转载或用于商业用途。

附录一的九张操作卡，授权教师、教研员与家长在非商业的教学与教研场合自由复印使用，复印时请保留出处。

内容提要：答案越来越便宜的时代，一张打了勾的卷子还能证明什么？本书提出一个判断数学学习是否真正发生的标准，并把它做成一张贯穿全书的格子：横轴问“这次做对了没有”，竖轴问“结构还在不在”。全书的注意力落在那个最容易被表扬、也最危险的位置——做得又快又对，可正确里面已经没有任何东西还能重新长出来。三编分别回答：什么才算真的学会；一堂课该留下什么；为什么它在真实课堂里这么少发生。书末附九张可复印的操作卡、一份阅读地图与一份验证方案。

本书承接：王德生《课堂学习发生学——一节课留下的那样东西，账本上没有它的字段》（德麦国际出版社，ISBN 978-1-970820-90-4）。

同著者：《课堂的智慧——课堂内生对象、运行条件与失效动力的可证伪理论纲领》（德麦国际出版社，ISBN 978-1-970820-95-9）。本书为其数学篇与读者版，可独立阅读。

适读人群：小学至初中数学教师、教研员，以及愿意认真陪孩子想一想的家长。

关键词：数学教育 | 学习评价 | 课堂观察 | 迁移 | 人工智能与学习

## 作者简介

### 阳涌

阳涌，英语语言文学硕士，拥有十六年一线教学与课程改革经验，长期致力于语言、数学、思维与人工智能教育的深度融合，曾指导多名学生进入世界一流高校。

自 2019 年起，他跟随王德生博士系统研修“三视角”“321 智慧”、SIO 及 SDE 本体论，形成以整体生成、差异推进、关系纠缠为核心的教育观。现专注于生成式学习、数学底层结构、创新阅读与教育智能体研发，构建“AI + 创新教育”体系，推动知识在学科、生活与真实问题之间迁移、通融与创造。

他的研究横跨教育评价、课程理论、创造力退化与 AI 时代的主体性，已发表论文四十余篇。贯穿其中的是同一个追问：**主体的能力与判断，是在什么条件下被生成，又是在什么条件下被悄然废弛。**从战争到课堂，他反复守护同一样东西——那种在效率与确定性的挤压下最先被牺牲、却恰恰是创造之根基的、必要的不确定性。

**已出版：**《课堂的智慧：课堂内生对象、运行条件与失效动力的可证伪理论纲领》（德麦国际出版社，260 页，约 20.5 万字，ISBN 978-1-970820-95-9）。

《答案之后》是作者的第二部专著，也是《课堂的智慧》的数学篇与读者版。

# 前言

## 一、十六年里反复遇到的那件事

我教了十六年书。有一件事我遇到过很多次，每次都不知道该拿它怎么办。

一个孩子，作业本干干净净，正确率一直很好，你完全有理由把他归到“学得不错”那一类。然后到了某个节点——分数进来的时候、负数进来的时候、字母进来的时候、第一次要求证明的时候——他突然就塌了。不是慢慢退步，是塌。

塌下来之后，通常的解释有两种。一种说是“新知识太难了”，于是补课、加练、再讲一遍。另一种说是“基础不牢”，于是往回补，把前面的题再刷一遍。

两种解释都很顺，也都做过。而我逐渐不信它们，理由很简单：**如果基础真的不牢，为什么之前那么长时间里，一次都没有露出来？**

一个真正不牢的地基，不该在三年里每一次检查都合格，然后在第四年的某一天整个垮掉。**除非我们一直在检查的那件事，根本不是地基。**

我印象最深的一次是一个初一的孩子。整个小学他的计算又快又准，几乎没错过。上了初中第一次遇到含字母的式子，他卡在一个地方—— $3a+2a$  他会，写成  $5a$ ； $3a+2b$  他也会，知道不能合并。可是当题目变成“一支笔  $a$  元，一本本子  $b$  元，买 3 支笔和 2 本本子一共多少钱”，他写下的是  $5ab$ 。

我问他怎么想的，他说：“都是买东西的钱，应该加在一起。”

**他不是不会合并同类项，他是从来没有把“同类”这件事当作一个需要判断的东西。**小学六年里，凡是要相加的量，题目从来没有让他判断过它们能不能相加——因为题目里出现的从来都是同一种东西。**那六年正确，一次都没有检验过这一条。**

这本书就是从这个不服气开始的。

## 二、答案的价格塌了

这个问题在今天变得更急。

十年前，一个孩子能稳定地把题做对，这件事本身就是稀缺的。我们有充分的理由把正确当作学习的证据——**因为得到正确太难了，能做到的人一定经历了什么。**

今天不是这样了。计算器、搜索引擎、解题软件和生成式人工智能，可以在几秒钟内给出步骤、图形、讲解和好几种解法。**答案的价格塌了。**

价格塌下来的东西，不能再当作证据。

这句话听上去像一句时评，落到课堂上却极其具体：两份作业摆在面前，答案一样对、步骤一样齐、字迹一样清楚，**我们凭什么判断哪一份里面还留着能继续生长的东西？**

我找过现成的答案。理解、迁移、灵活、深度、素养——这些词都指出了方向，可是没有一个能帮我在那两份作业之间作出判断。它们都需要先假定一个人“懂了”，再去描述那种懂的性质。而我需要的恰恰相反：**一个不依赖“你觉得他懂不懂”的问法。**

## 三、账本上没有的那几栏

这本书不是从我一个人开始的。它的来路要交代清楚。

王德生老师有一部《课堂学习发生学》，全书立在一句话上：

课堂里决定后果的那些东西，有一大部分是课堂自己造出来的；而课堂的账本只为“不是它造的”那些东西设了字段。于是它每天都在生产它看不见的东西，并把看不见的部分算作零。

这句话第一次读到的时候，我在办公室里坐了很久——因为它回答了我前面那个不服气。

那本书开头讲了三间教室。第一间是数学教室：学生刚做完二十道同类型的题，正确率接近百分之百；老师把题里一个常量改成变量，删掉那句提示，教室安静下来——学生并没有忘记公式，他们不知道现在该把什么当作问题。第二间是工程教室，学生按流程做完了方案，评分表每一项都达标；换掉使用者和预算，撤走量表，他们仍会操作软件，却把时间花在打听“老师到底要哪个答案”。第三间是艺术教室，一次讲评让学生很受触动，他能复述老师的每一句话；两周后面对一件新作品，他仍在等某个权威先说“该怎么看”。

三间教室有一个共同点，而它不在学生身上，也不在老师身上：

**在事情正在发生的那一刻，三间教室的账本都是满分的。**

每一栏都填得漂亮，每一栏都是真的。出问题的不是记账的诚实，是账本上没有的那几栏。

那本书还有一句更不留情面的，是我这本书里所有“越好越糟”判断的来源：**账本每完备一分，那一段就被压掉一分。**课程建设越完备，班级自己造出来的东西越少；老师在一门课上教得越久，他给出的“下一步”越稳定唯一，于是越容易把本来有好几个答案的地方讲成只有一个。**这三处都不是失职的产物，是尽职的产物。**

那本书里还有一句最硬的，我用了很久才敢说自己读懂：**课堂真正留下的那样东西无人持有，因而不能在任何一个时点对任何一个人结清。**而账本恰恰是按“持有人+结算时点”造出来的——**两者在这里正面顶上。**

它接着指出三处熄火的位置（入口、伴生物、出口），并且证明三处的效果**相乘而不相加**——单修任何一项，收益的上界由另外两项封死。这几件事的完整交代在正文的理论章里，那里有原书的表和数字。

#### 四、三本书，一条线

我自己的上一本《课堂的智慧》，接着这条线又走了一段。它处理的是：**课堂用来宣布“学会了”的五类凭据——正确、熟练、记住、完成、作品像样——全都可以在真正的学习没有发生时出现，而且往往出现得更漂亮。**

那本书九章各找到一种最难被识别的失败形态，最后并成一张表。放在一起之后显露出一件写作时并未预设的事：**九个靶格没有一个是明显的坏课堂，它们全都在被评价的那个维度上合格。**

现在你手上这一本，是这条线的第三段。

- 《课堂学习发生学》问：账本上少了哪几栏？
- 《课堂的智慧》问：那么失败会长成哪九种样子，为什么它们长得像成功？
- 《答案之后》问：那么这道题，明天该怎么问？

三本可以分开读。你手上这一本不需要先读前两本——它们的骨架，我整章搬到了正文的理论章里，大约一万两千字，是全书唯一一处不谈具体题目的地方。

## 五、为什么偏偏是数学

有人问过我：这套判断既然对课堂普遍有效，为什么要专门写一本数学的？

因为数学是这件事最容易发生、也最容易被掩盖的一门学科。

**数学的答案是唯一的。**语文的作文没有标准答案，一份作业好不好，老师必须读进去才知道；而一道数学题只要对上答案，判断就完成了——**判断可以完全不进入过程。**这一条本来是数学的优点（客观、可比较、可大规模检验），但它同时意味着：数学是唯一一门可以在完全不了解学生怎么想的情况下，把整套评价跑完的学科。

**数学的步骤是可以脱离理由存在的。**"移项变号""除以分数等于乘倒数""看见比就写比例式"——这些动作可以完全正确地执行几百次，而支撑它们的关系一次都没有出现。**别的学科很难做到这一点：**一个孩子如果不理解一段史料，他复述出来的东西一眼就能看出是背的。

**数学的内容是层层叠上去的。**所以前面欠的账不会立刻爆，会一直往后压，压到分数、负数、字母、函数、证明这些节点上一次性爆掉——这就是我在第一节里说的那种"塌"。

**还有第四条，是最近几年才变得要紧的。**人工智能最先攻下的就是数学。写作、绘画、翻译都还在争论机器做得好不好，而一道中学数学题，机器已经能在几秒钟里给出正确答案、完整步骤和三种解法，而且讲得比大多数人清楚。**在所有学科里，数学是"答案最先变得不值钱"的那一门。**

这意味着数学老师是第一批被逼到墙角的人——**也意味着这门学科最先需要一套新的判断办法。**别的学科还有几年时间，数学没有。

三条合起来：**数学是最容易让"死的正确"活得又久又体面的学科。**所以它值得单独写一本。

## 六、这本书里有什么：九句话

正文九章，每一章都在还一句话的债。把九句话放在一起，就是这本书的全部内容：

### 第一编·什么才算真的学会

1. 做对一道题，可能发生两件完全不同的事：一件是路走通了；另一件是路走通了，而其他没走的路全被封死了。
2. 一次成功不等于一次能力——先查清楚：这个答案，是由孩子的哪一条关系撑起来的？
3. 新知识学进去了不叫学会。**真正的学会，是新东西反过来把旧东西改写了一遍。**

### 第二编·一堂课该留下什么

1. 一堂课上得好不好，不看这堂课结束时孩子会了什么；**看下一堂课能不能绕开它。**
2. 学习不是把路走对，是走偏的时候还知道自己在哪儿。
3. 一个概念真正的边界，不只由"哪些例子该用它"构成，**也由"哪些几乎一样的例子必须忍住不用"共同构成。**

### 第三编·为什么它这么少发生

1. 数学学习很少发生，不是因为错误太多，而是因为**课堂过早地生产了正确。**
2. 课堂被安排得太满了——满到没有一处余地，可以让一个还没算清楚的差异活到下一轮。
3. 便于搬运的那个单位，取得了定义学习单位的权力。**知识点是一个运输单位，不是一个学习单位。**

九章之外，还有一章理论、一篇结论、三个附录（九张可复印的操作卡、一份阅读地图、一份给研究者的验证方案），以及十二张图。

如果你现在只想验证这本书值不值得读，请翻到第七章那一节《一次真算》。那一节里的数字，是这本书能不能站住的关键。

## 七、这本书押在哪三处

**第一处，判断的位置。**全书所有的分界，没有一处落在当堂——全部落在后面：换掉承重条件之后、撤掉支架之后、材料被替换之后、走偏之后、遇到近邻之后、隔一段时间之后。这不是我偏爱延迟测量，是这套判断的对象性质决定的：**凡是以“参加未来经验”为存在条件的东西，就不可能在事情结束的那一刻被完整验收。**

**第二处，那三条把课堂往回拉的力。**第三编三章讲了三个自我加固的循环，它们有一个共同点，也是这本书最不好受的判断：**三条力都随着“做得更好”而加强，不是随着“做得更差”而加强。**也就是说：**改进本身可以是病因。**这句话在第三编三章里各出现一次形态，理论章第九节把三种形态并在一起。

**第三处，一个真算出来的数字。**第七章有一节用的是一批带误解标注的真实题库：1869 道题、1604 种误解。结果是——同一个知识点上任取两种误解、任取一道该知识点的题，**只有百分之七的情况下，这道题能让两个持不同误解的孩子给出不同的答案。**这个数字把“不可识别”从一个说法变成了一件可以查的事实。

## 八、这本书为什么长这个样子

有几个体例上的决定，我在这里交代一下，免得你读到一半觉得奇怪。

**每一章都从一个课堂场面开始，不从道理开始。**这不是为了好读。教育书里最常见的一种失效是：道理全对，落到具体一节课上却不知道该改哪一句话。**所以我强迫自己每一章先写出一个能看见的场面，写不出来的判断，就说明我自己还没想清楚。**

**每一章都从三门跟数学教育无关的学问进入。**你会遇到植物学、考古学、绘画、密码学、免疫学、排队论、审计学……九章一共二十七门。它们**不是比喻**——理由和做法，理论章里有完整的交代。你现在只需要知道一件事：**读不懂的地方是我的责任，不是你的；**而且每一章的第三节整节跳过，后面照样读得懂、照样能用。

**每一章的最后一节都叫“这套眼光的分寸”。**那一节讲的是这一章什么时候不该用。我把它固定成体例，是因为一套判断工具最危险的时刻，恰恰是它开始被到处使用的时候。

**全书只用一张格子。**九章会给出九个不同的名字，但它们全都挂在同一张格子上。**记住那张格子，九个名字忘掉也不要紧。**

## 九、一件我改过的事

初稿里，第五章有一句是“偏差出现的时候，给他十秒钟”。

后来我去查了——这件事在五十年前就被人拿秒表量过。实测的数字是：老师提完问题，**平均等不到一秒就会开口**；学生答完之后到老师开口，平均 0.9 秒。而当这两个停顿被延长到三到五秒，变化是成组出现的：回答变长、答不上来的次数减少、开始出现带猜测性质的回答、学生主动提问增加，**而且平时被认为反应慢的孩子，发言次数明显增加。**

于是我把那句话改了：**不是十秒，是三秒。三秒是有实测支撑的门槛，十秒是野心。**

我把这件小事写进前言，是因为它是这本书的一个样本。**书里每一处判断，我都是先写下自己的看法，再去查、去撞、去改的。**有几处改得比这一处更狠——比如那三门学问从我草稿里删掉的三句漂亮话，正文里都写明了是哪三句。

顺带说一句：三秒之所以难，正是因为一秒是常态。**三秒是常态的三倍。**

## 十、这本书打算怎样被推翻

一本书如果只能被同意，不能被推翻，那它就还不是一份可以拿来用的判断。所以我把这本书最可能错的地方写在前面。

**如果这些做法只是“好老师本来就会做的事”，那这本书就没有必要。**判决办法很简单：控制住老师的经验与班级基础之后，那几项记录是否还能预测后来的表现。若不能，这本书应当退回一份经验总结。

**如果“当堂看不出来”这条不成立，全书的地基就塌了。**只要能证明存在一套当堂就能采集、并且能稳定区分那九种情形的指标，这本书的第一条推论立刻作废——**而我会很高兴，因为那意味着这件事比我想的容易得多。**

**如果第七章那个百分之七在中文题库上算出来是百分之七十，第七章第一阶段就不成立。**附录三写清了怎么复算，用的是一套任何人都能拿到的公开数据。

我把这些写在前言，是因为这本书要求老师改变的东西虽然小（三秒、一句话、一道近邻题），但它要求你相信一件反直觉的事：**你的课上得越好，某些东西越难被看见。**一个提出这种要求的人，有义务先说清楚自己会在哪里输。

## 十一、三个请求

**第一，别把它当成一套教学法。**这本书给你的是一双眼睛，不是一张课表。眼睛和课表的差别，我在导论里会说清楚。

**第二，别拿它给孩子分类。**书里所有说法讲的都是“这一道题上、这一次的结构位置”，不是“这个孩子是哪种人”。一切能用来给人分类的东西，都会被人拿来分类——这是我唯一真正担心的误用。

**第三，用完之后，回来告诉我哪里不对。**尤其是哪里“学生根本不会这么说”。这本书的案例大多是我构想的典型场面，每一处都已标明；它们能把结构说清楚，但它们不是实录。**第二版和第一版的差别，全在你的教室里。**

如果这本书最后只留下一句话，我希望是这一句。它不需要任何术语，任何人都能在明天早上用：

**让孩子先说，你再说。**

阳涌

## 导读·这本书怎么读

这本书有九章，每一章都可以单独读，但它们之间有顺序。下面给三种读者三条路径，你可以从最急的那一处开始。

### 全书的地图

| 编   | 问题        | 章     |
|-----|-----------|-------|
| 第一编 | 什么才算真的学会  | 第一至三章 |
| 第二编 | 一堂课该留下什么  | 第四至六章 |
| 第三编 | 为什么它这么少发生 | 第七至九章 |

第一编讲**要看什么**，第二编讲**怎么让它长出来**，第三编讲**为什么它长不出来**。

三编的关系不是并列。如果你只读第一编，你会得到一双更好的眼睛，但可能会觉得“道理都对，做不到”；如果你跳过第一编直接读第二编，那些方法会退化成新的教学花样；而第三编解释了为什么前两编在真实的学校里这么难落地——**读完第三编再回头看第二编**，你会知道该在哪里用力，也会知道哪些事不该由你一个人扛。

### 三条路径

**如果你是老师，时间很紧：**

先读**导论**，再读**第七章**，然后翻**附录一的卡七**。（**理论章可以先跳过**——它是这本书的骨架，来自王德生老师的《课堂学习发生学》和我的上一部《课堂的智慧》；什么时候想知道“这些判断凭什么”，再回来读。）这三处加起来不到一小时，能立刻改变你明天的一句话。有余力再读第五、六章——这两章对备课的改动最直接。

**如果你是教研员或者学校管理者：**

先读**理论章**——那一章的第七节讲三层不能互相还原，正是你的位置所在。然后读**第三编**（第七、八、九章），再回头读**导论和结论**。这本书里真正需要你的权限才能改的东西，都在这三章和结论的最后一节。

**如果你是家长：**

读**导论、第一章和结论**。中间那六章是给课堂设计用的，你不必读完；但第一章那张记录表和结论里给家长的那一句话，值得贴在书桌上。

### 每一章的结构都一样

九章用的是同一个模子，你读完第一章就会熟悉它：

1. **一个课堂现场**——具体到某道题、某句话；
2. **一句判断**——加粗，全章都在还这笔债；
3. **三门学问怎么看**——它们打架的那一点；
4. **一个说法和一张格子**；
5. **几个数学场面**——从小学到初中，这一部分是全书最实用的血肉；
6. **一张操作卡**——明天就能用；

## 7. 这套眼光的分寸——什么时候不该用它。

如果你时间很紧，**每章直接跳到第五、六两节**也完全读得懂。前面四节是为了让你信服，后面两节是为了让你能用。

## 关于书里那二十七门学问

翻开目录你会看到植物学、考古学、密码学、免疫学、排队论、行为生态学、审计学、系统分类学……九章一共请了二十七门学问出场。

**我要先打消一个顾虑：你不需要先懂它们。**

每一门学问出场的时候，都从零讲起——它在研究什么、拿什么现象说事、关键的那个词是什么意思、用一个日常生活里见得到的例子。**读到看不懂的地方，是我的责任，不是你的。**每一章的第三节末尾还有一张小表，把那一章的专业词收成一行一句的解释，可以随时翻回去对。

也请你不要把它们当成比喻。这本书从头到尾没有说“学习就像一棵树”“课堂就像流水线”——那类句子听着舒服，但什么也判断不了。

**它们出场的真正理由是这个：**每一组三门学问，都各自认真回答过一个和课堂同构的问题，**而且给出了互不相容的答案。**

举个例子。第一章问的是“没有实现出来的东西算不算存在”：

- 植物学说**算**——一个没长出来的芽仍然保有再长出来的能力；
- 考古学说**要看痕迹**——但它同时警告你，一样的痕迹可能来自不一样的过程；
- 绘画说**算，但要看它有没有参与选择**——被覆盖的构图塑造了最终的构图，可改得多不等于想得深。

三家吵起来，我才被迫把话说得比原来精确。如果三家意见一致，那这一章就不用写了。

还有一件更要紧的事：**这些学问经常反过来限制我。**植物学不让我说“每个孩子都有无限潜能”——因为芽的数量有限，让所有芽同时萌发的植物会变脆弱；考古学不让我说“把过程记全了就能判断学习”——因为痕迹会骗人；绘画不让我说“多改几遍就是想得深”。

**这三句话原本都在我的草稿里，是它们把这三句删掉的。**书里每一章都有这样一处，我会如实写出来。

最后，如果你实在对某一门学问没兴趣，**每章的第三节可以整节跳过。**跳过之后第四节的判断照样读得懂，第五、六节照样能用。三门学问是我用来把话说准的工具，不是读者必须缴的学费。

## 一件请你留意的事

书里的课堂场景，多数是我构想的典型场面，不是某一节课的实录。凡是构想的场面，段首都写明了“这是我构想的场景”。

我这样做，是因为构想能把结构说得比实录更清楚——真实课堂里的一节课，往往同时混着五六件事。但请你在使用的时候，把它当作一张需要你亲自去验证的图纸，而不是一份已经验收过的报告。**你在自己教室里看到的，永远比这本书里写的更准确。**

## 三个附录

**附录一**是九张操作卡，可以复印。

**附录二**是一张阅读地图，告诉你这些想法从哪儿来，以及可以往哪儿继续读。

**附录三**写给愿意把这些说法做成研究的人。

如果你只想拿走一样东西，拿走附录一的第七张卡。上面只有一句话。

# 目录

|                           |           |
|---------------------------|-----------|
| 作者简介.....                 | 3         |
| 前言.....                   | 4         |
| 导读·这本书怎么读.....            | 9         |
| 图目录.....                  | 13        |
| 导论·答案变便宜之后.....           | 14        |
| 理论章·凭据、九格与三层.....         | 18        |
| <b>第一编 什么才算真的学会.....</b>  | <b>32</b> |
| 第一章·没走的那条路还在不在.....       | 33        |
| 第二章·这个答案，是谁在替他撑着.....     | 43        |
| 第三章·旧知识会不会被改写.....        | 52        |
| <b>第二编 一堂课该留下什么.....</b>  | <b>62</b> |
| 第四章·留下下一堂课绕不开的东西.....     | 63        |
| 第五章·第一次走偏的时候.....         | 72        |
| 第六章·对“几乎一样”说不.....        | 82        |
| <b>第三编 为什么它这么少发生.....</b> | <b>91</b> |
| 第七章·正确来得太早.....           | 92        |
| 第八章·课堂越高效，越难发生.....       | 101       |
| 第九章·越容易搬运，越难长出来.....      | 110       |
| 结论·给三种人的三句话.....          | 118       |
| 后记·这本书是怎么写出来的.....        | 121       |
| 参考文献.....                 | 123       |
| 附录一·九张操作卡.....            | 130       |
| 附录二·这些想法从哪儿来.....         | 134       |
| 附录三·给研究者：怎样验证这些判断.....    | 138       |
| 名词索引.....                 | 142       |

## 图目录

|      |                     |     |
|------|---------------------|-----|
| 图 1  | 全书唯一的工具：一张格子.....   | 15  |
| 图 2  | 两株外形相同的树.....       | 38  |
| 图 3  | 一次正确的载荷路径：五根柱子..... | 48  |
| 图 4  | 代谢的五个动作.....        | 58  |
| 图 5  | 关系栖造：终点换手.....      | 68  |
| 图 6  | 五步校航，与走偏之后的那一问..... | 77  |
| 图 7  | 六步校界.....           | 87  |
| 图 8  | 远的正例与近的负例.....      | 88  |
| 图 9  | 正确性消音循环.....        | 97  |
| 图 10 | 利用率与等待时间.....       | 102 |
| 图 11 | 余量清零循环.....         | 106 |
| 图 12 | 知识集装箱循环.....        | 115 |

## 导论·答案变便宜之后

### 一张打了勾的卷子

一张数学卷子摆在桌上。

答案正确，步骤完整，书写整洁。老师在末尾画了一个鲜红的勾，家长松了一口气，孩子自己也相信：这部分我学会了。

我想请你和我一起，在这个让人安心的时刻停一下，问一句不太礼貌的话——

**这张卷子究竟证明了什么？**

它证明了这个孩子，在这个时间、这道题、这套说法、这组条件之下，完成了一次被认可的操作。

它没有证明孩子知道自己把什么当成了对象；没有证明他为什么选这个单位；没有证明题目里的几个量凭什么能建立这种关系。更没有证明的是：当题目换一种画法、删掉一个熟悉的条件、把“每份多少”改成“多少为一份”的时候，他还能不能重新把问题组织起来。

过去，这个缺口很容易被忽略。因为得到答案本身就很难。一个孩子能稳定地算对，已经是一种稀缺的本事，我们没有余力再去追问后面那些。

今天不一样了。

计算器、搜索引擎、解题软件、生成式人工智能，可以在几秒钟里给出步骤、图形、讲解，甚至好几种解法。**答案越容易拿到，答案就越不能单独充当“学习已经发生”的证据。**

这不是说答案不重要。没有结果，学习很容易退化成不受约束的自我感动：讨论得很热烈，谁也说不清最后到底成立了什么。问题在别处——我们长期把“学习留下的一个结果”，误认成了“学习本身”。

这本书，就是想把这两件事分开。

### 两个孩子都做对了

要分开它们，教育里那些熟悉的词好像还不太够用。

理解、迁移、灵活、深度、素养——每一个都指出了方向，可是当两份作业摆在你面前，答案一样对、步骤一样齐，这些词没有一个能帮你判断：**哪一份里面还留着能继续生长的东西，哪一份的正确已经把所有没走的路都封死了。**

这就是全书要回答的问题。它不是“怎样教得更好”，也不是“怎样让学生更喜欢数学”——那些都是好问题，但都得等这一问先有答案：

**一个孩子做对了。凭什么判断，学习真的在他身上发生了？**

我把这本书写成三编，其实就是三个连着的问题：

**第一编问：什么才算真的学会。**（如果只看答案不够，那要看什么）

**第二编问：一堂课该留下什么。**（要看的東西，怎么才能长出来）

**第三编问：为什么它这么少发生。**（为什么大部分课堂长不出来，而且越努力越长不出来）

### 一张格子

这本书从头到尾只用一件工具。它是一张格子，两条轴。

横着看，问的是**这一次做对了没有**——就是我们平时看的那件事：答案对不对，步骤全不全，速度快不快。

竖着看，问的是**结构还在不在**——如果我把这道题的一个关键条件换掉，让原来那条路走不通了，这个孩子还能不能重新决定：现在研究的是什么、以什么为一份、哪两个量之间该建立关系、原来那个方法是从哪一步开始不管用的。

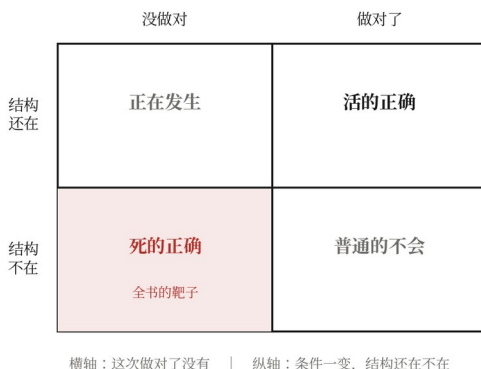
两条轴一交叉，得到四个格：

|      | 这次做对了 | 这次没做对 |
|------|-------|-------|
| 结构还在 | 活的正确  | 正在发生  |
| 结构不在 | 死的正确  | 普通的不会 |

右下角，"普通的不会"，我们都认得，也知道该怎么办。

左上角，"活的正确"，是我们想要的。

**这本书全部的注意力，在左下角。**



横轴：这次做对了没有 | 纵轴：条件一变，结构还在不在

图1 全书唯一的工具：一张格子。九章各回来一次。

一个孩子做得又快又对，本子干净，从不掉队，只是——他的正确里面，已经没有任何东西还能重新长出来了。我把这一格叫作**死的正确**。

它不是错误。它往往比"活的正确"看起来更漂亮：更快、更整齐、在熟悉的题型里得分更高。它也不是孩子的错，更不是老师的错。它是一种位置，一种状态，任何一个孩子在任何一个内容上都可能落进去，今天在这一格，明天在那一格。

我之所以要花一整本书对付它，是因为它有三个性子：

**第一，它在短期指标上表现为变好。**题做得更快，错得更少，作业更整齐，家长更放心。

**第二，它失效的时候不报警。**只要题目保持熟悉，孩子会一直成功。直到某一次结构真的变了——换个画法、少个条件、进了下一个学段——才突然塌下来。而那个时候，我们通常会把原因归到"这次题难"或者"他最近不用功"。

**第三，它会自我加固。**孩子越依赖标准步骤，成绩越稳定；成绩越稳定，老师越确信这套教法有效；越确信，就越舍不得把时间花在试探、比较和解释上。**最后是成功本身，挡住了学习继续发生。**

这三句话，请你先记下来。后面九章，你会在九个完全不同的场景里，一次又一次认出同一张脸：一年级的退位减法里，初中的辅助线里，一堂被评为优质课的公开课里，一个进度条永远显示“已掌握”的学习平台里。**它们是同一个东西的九张脸。**

到第三编，我会把这件事挑明：为什么一个越来越好的系统，会越来越难长出学习。那是这本书最不好受的部分，也是我认为最要紧的部分。

## 关于这本书里的“三门学问”

**在往下说之前，先讲清楚一件事：这本书不预设你懂这些学问。**

每一门都从零讲起。腋芽是什么、集装箱标准改变了什么、免疫耐受为什么不是“没认出来”、一个统计学家说“这个观测不可识别”到底在说什么——全都用日常能见到的例子讲，讲不明白算我的。每一章的第三节末尾还有一张小表，把那一章的词收成一行一句。

你会发现每一章都从三门和数学教育没什么关系的学问开始：植物学、考古学、绘画；物理、心理、工程；诊断、法律、航海；音乐、密码、免疫；排队论、动物行为、审计……

我先说清楚这是在干什么，免得读起来像卖弄。

这不是比喻。我不是想说“学习就像一棵树”“课堂就像一条流水线”——那种句子听着舒服，但什么也不能判断。

我请这些学问出场，是因为它们各自都在自己的领域里，**认真回答过一个和数学教育同构的问题，而且给出了互不相容的答案。**植物学问过“没长出来的枝条算不算存在”；考古学问过“消失了的过程能不能从废墟里重建”；绘画问过“被覆盖掉的线条还算不算作品的一部分”。三家答得不一樣，正因为不一样，它们逼着我把话说得比原来精确。

而且，它们经常反过来限制我。植物学不允许我说“孩子有无限潜能”——芽的数量有限，位置有限，养分有限，让所有芽同时萌发的植物会变得脆弱。这一条直接砍掉了我原本很想写的一句漂亮话。每一章都有这样一处：**某门学问反过来削弱了我的主张。我会在正文里如实写出来。**

## 关于这本书里的例子

这本书里的课堂场景，多数是我构想的典型场面，不是某一节课的实录。我构想它们，是因为构想能把结构说得比实录更清楚——真实课堂里的一节课，往往同时混着五六件事。

所以我立了两条规矩，全书通行：凡是构想的场面，段首写明“这是我构想的场景”；凡是真实的课堂材料，写清年级、时间和课的类型。孩子说的话如果是我替他说的，不打引号冒充原话。

这两条规矩，是这本书对你的基本尊重。

## 这本书不做的三件事

**它不提供一套教学法。**这本书不会告诉你每节课的前五分钟该干什么。它给你的是一双眼睛，不是一张课表。眼睛比课表管用得多——课表只在它被设计的那种课上有效，眼睛到哪儿都能用。

**它不反对讲授、练习和自动化。**这一点我要说重些，因为很容易被误读。数事实、基本运算、符号书写，需要练到不用想。没有稳定的主干，什么也长不出来。这本书反对的从来不是主干，而是**把主干的稳定，误认成整棵树已经长成。**

**它不用来给孩子贴标签。**这本书里所有的说法，说的都是“这一道题上、这一次的结构位置”，不是“这个孩子是哪种人”。同一个孩子，可能在分数上枝繁叶茂，在几何上封得死死的；今天没长出来，不等于以后不会。**一切能用来给人分类的东西，都会被人拿来分类**——这是全书唯一需要你警惕的误用，后面我会再说一次。

## 那句话

好了，让我把这本书想说的那句话先放在这儿。你可以现在读一遍，读完全书再回来读一遍：

**真正的会，不是从此不再迷路。真正的会，是旧地图失效之后，还能重新长出路来。**

现在，我们从一张满分的卷子开始。

## 理论章·凭据、九格与三层

承接《课堂的智慧》

### 一、为什么这本书要先搬一章理论过来

这本书讲数学。但它用的那双眼睛不是从数学教育里长出来的，是从上一本书里长出来的。

《课堂的智慧》处理的是一个更一般的问题：**课堂每天都在宣布学习已经发生，凭据是正确、熟练、记住、完成、作品像样。这五类凭据全都可以在真正的学习没有发生时出现，而且往往出现得更漂亮。**

那本书是一份理论纲领，九章各找到一种最难被识别的课堂失败形态。这一本把同一副眼光收进数学——收进分数单位、三角形的高、移项、正比例这些每天都在发生的地方。

问题是：如果你没读过那一本，这本书里的很多判断会像是凭空冒出来的。为什么“课堂看不出来”？为什么“分界全落在后面”？为什么第三编那三个循环长得那么像？

**所以我把上一本的骨架整章搬到这里。**读完这一章，你不必先去读那本书，就能知道这本书的每一个判断站在什么上面。已经读过那本书的人，可以直接跳到下一章——这一章对你没有新东西，只有一张新的对照表在最后。

**这一章是全书唯一一处不谈具体数学题的地方。**它讲的是骨头。

### 二、更上游的那一本：账本与它没有设的字段

这套眼光还有一个更上游的来源，必须交代：王德生老师的《课堂学习发生学》。

那本书的母题，我在前言里已经整句引过——**课堂每天都在生产它看不见的东西，并把看不见的部分算作零。**这一节把它底下的四层结构拆开。

**第一层：课堂在生产它自己的条件。**

那本书的第二编说的是——课堂里的**终点、次序与条件**，各有一部分是这个过程自己临时造出来的：

- **终点会易主。**一个手段可以在中途临时取得终点的资格（“我们先把这个图画准”），然后再交还回去。这一段被那本书叫作**换位区间**。
- **“下一步该做什么”不一定只有一个答案。**从目标往回倒推，解可以是一个、多个，或者没有。那本书叫它**可倒推区间**，并给了一条与通行做法相反的建议：**档位不确定的时候，默认推进而不是默认倒推**——因为在唯一区误用推进，损失是有界的效率损失；在多解区误用倒推，损失是能力损失，**而且在原来的场合里完全看不见**。
- **条件有一部分是自造的。**一个班在学习过程中攒下的记号、约定、工具、错误清单，就是**自建条件**。而这里有一个让人不舒服的推论：**一门课程越“完备”，班级自己造出来的条件越少**——完备度与自建存量成反比，而完备的每一步在任何评估里都记为进步。

三条合起来是同一句话在三个部件上的形态：**次序里被当作外生的那些量，有一部分是内生的。**

**第二层：那样东西无人持有。**

前言引过那句“它无人持有，因而不能在任何一个时点对任何一个人结清”。它的后半句同样要紧：它靠被交出去、被改写、被继续使用而继续存在；一旦不再被交出去，它不是被遗忘，是停止存在。

账本恰恰是按“持有人+结算时点”造出来的。定义与账本在这里正面顶上——这就是为什么这套眼光里所有的判断都不肯落在当堂。

它给了四条构成条件，这本书九章各用到一条或几条：**扰动留痕**（受扰之后留下痕迹而不是复原）、**撤架再入**（支架撤掉还能重新进去）、**异境生差**（换情境仍能长出差异）、**持位可换**（核心功能不锁死在某位老师或某个尖子生身上）。

还有一条提醒，直接决定了这本书所有记录表的做法：**这样东西平时不显露，只在一次新的学习被重新发起时露一次面。而传统测验只查旧答案能不能被直接提取，因此系统性地**把“休眠”读成了“遗忘”。

**第三层：三处熄火，相乘不相加。**

那本书的第三编说，一次改变会在三个位置各熄一次火——**入口、伴生物、出口：**

| 位置  | 熄火形态                                       | 它的靶格是什么样的课       |
|-----|--------------------------------------------|------------------|
| 入口  | <b>空窗：</b> 上一轮打开的感受期有时限，窗开着而无人到达           | 课程表执行完美、进度分毫不差   |
| 伴生物 | <b>悬置连接：</b> 为讲清一处而挡住其余，被挡住的部分在暗处结成无人拆除的关联 | 讲解清晰、每处都先“暂时不考虑” |
| 出口  | <b>冻结构型：</b> 还在重组的时候被读了数，并按这个读数配置了此后半年     | 评价及时、反馈频密、分层精准   |

**请看最右边那一列：三处的靶格都是好课。**课程表执行到位、讲解层层铺垫、评价及时精准——这三样正是教学管理最想要的东西，而它们恰好分别是三处熄火最容易发生的条件。

三章合起来给出一条任何一章单独都推不出的判断，也是那本书最锋利的一条：

**三处的效果不是可加的贡献量，而是串联的通过率——它们相乘，不相加。**

推论很硬：**单修任何一项，收益的上界由另外两项封死。**若入口的通过率只有 0.2，你把中间那项从 0.8 提到 0.95，全部收益不超过  $0.2 \times 0.15$ ——而这笔投入在任何一份改革方案里都会显示为“评价体系显著优化”。

**第四层：为什么它们不报警。**

那本书对三处熄火还有一个共同的判断，我认为它比三个对象本身更值钱：

**它们在发生的时候，不产生任何可以被现有记录系统接收的事件。**

窗口关上时什么也没发生；悬置连接在原型里不产生任何差错；被过早读出的那个数在所有可见指标上都正常，因为后续条件正是按它配置的，于是它自我实现。

**所以它们不是被漏记的，是不可能被记的。**现有记录的每一栏都需要一个事件来触发，而这三处的定义特征恰恰是不触发事件。

**看不见的部分因此被算作零，而不是被算作未知。未知会引出调查，零不会。**

由此还得到一条对策上的转向，这本书第三编的落点正是从这里来的：**要观察那些不产生事件的位置，必须主动去看，而且要在没有信号的时候去看。**

**这三本书的关系，可以放在一起看：**

|                | 问的是          | 落点                   |
|----------------|--------------|----------------------|
| 《课堂学习发生学》（王德生） | 账本上少了哪几栏     | 课堂自己生产它看不见的东西，并把它算作零 |
| 《课堂的智慧》（本书作者）  | 那么失败会长成哪九种样子 | 真正的失败形态恰好长得像成功       |
| 《答案之后》（本书）     | 那么这道题明天该怎么问  | 一张格子、九个名字、九张操作卡      |

下面这一章的其余部分，讲的是中间那一本。上游那本的东西，你已经在上面这一节里拿到了骨架。

### 三、三个问题必须一起问

课堂研究已经很拥挤。教学法、评价制度、课程设计、学习科学、教师专业发展，各自都有大量成熟的成果。再增加一份“好课堂应当具备哪些特征”的清单，不会让任何一位老师明天上课时多知道一件事。

所以上一本书只问三个问题，而且要求它们**必须一起被问**：

**第一问：何谓真正的学习。**这是“是什么”的问题。它追问的不是学习有哪些表现，而是：当我们说一个人真正学会了，究竟有什么此前不存在的东西开始存在？

**第二问：课堂如何运行，学习才会真正发生。**这是“怎样”的问题。它追问的不是好课堂有哪些品质，而是：课堂事件按怎样的先后关系发生，前一轮的结果怎样进入后一轮，哪一个环节失败会使整个过程退化？

**第三问：课堂为何很少发生学习。**这是“为什么”的问题。它追问的不是有哪些不利因素，而是：课堂系统中哪些矛盾会持续驱动行为改变，并把改变结果写回下一轮，使同一种失败不断复制？

**三问的次序不能颠倒，也不能只取其。**

只问第一问，会得到一份关于理想学习的描述。它可能很动人，却无法说明这种学习需要怎样的课堂条件，也无法解释为什么这些条件在真实学校里如此罕见。

只问第二问，会得到一套教学建议。它可能很实用，却不知道自己在服务于什么对象——**如果“真正的学习”没有被界定清楚，任何一种课堂运行方式都可以宣称自己有效。**

只问第三问，会得到一份对学校制度的批评。它可能很尖锐，却容易滑向道德指控：老师不够专业，学校太功利，考试压力太大。而恰恰相反：**那三条力都不是外部敌人，它们分别保护着学校不可缺少的东西——可问责、活动的连续、公共的稳定。真正困难的不是消灭它们，而是使它们不再垄断课堂。**

三问必须一起问，还有一个更硬的理由。**只有把对象、运行与动力放在一起，这套说法才有可能被推翻。**单独一问几乎总能自圆其说：学习没发生，可以说对象没界定好，也可以说运行方式不对，还可以说制度压力太大。三问同时给出判据，才可能出现“这三条互相矛盾”的情形——**而那正是一个理论可以被判错的地方。**

你手上这本书的三编，就是这三问在数学里的落地：第一编问什么才算真的学会，第二编问一堂课该留下什么，第三编问为什么它这么少发生。

## 为什么要请二十七门不相干的学问

上一本书九章，每章从三门彼此不相容的学科出发，合计二十七个学科入口。这本书沿用了同一个做法，你已经在每一章的第三节见过。

这件事最容易被误解成博学的展示，所以要把理由说死。

**第一，为什么是三门，而不是一门。**一门学科给出的是一个答案，两门给出的是一次比较，只有三门才可能出现“三家互相取消”的局面——而那正是需要的东西。如果三家能温和地相加，那这一章就没有必要写：把三段综述并排放好，读者拿到的是一份跨学科学习单，不是一个新判断。

**第二，为什么必须不相容。**找三门互相支持的学科很容易，也毫无用处——它们支持的多半是一个已经被广泛接受的说法。要的是三家对同一个问题给出无法同时成立的答案，然后追问：它们为什么不能相加？三家共同默认了什么，才使得各自的答案在自己的领域里都成立？

**推翻那条共同默认的材料，必须来自三家自己，不能从外面再搬一门学科进来——那样只会得到第四种立场。**

**第三，怎么判断一门学科是不是真的在承重。**上一本书用的是一条硬检验，这本书每一章也都过了这一关：

**把这门学科整条删掉，本章的判断、指标与预测是否仍然成立？若仍然成立，这门学科就只是修辞，应当删掉。**

这条检验听起来简单，做起来很不客气。它意味着一章里如果有一门学科只是提供了一个好听的比喻，它就该被拿掉——哪怕那个比喻是全章最漂亮的一句话。

**第四，学科反过来限制作者。**这是最容易被跳过、却最要紧的一条。请三门学科出场，不是为了让它们支持你，而是为了让它们有机会否定你。上一本书每一章都有这样一处；这本书也是——植物学删掉了“每个孩子都有无限潜能”，考古学删掉了“记全过程就能判断学习”，绘画删掉了“多改几遍就是想得深”。

一个跨学科的判断如果从来没有被它请来的任何一门学科削弱过，那多半说明它们只是被请来站台的。

## 四、五类凭据，全都可以在学习没发生时出现

课堂用来宣布“学会了”的凭据，归拢起来是五类：

**正确。**答案对。这是最硬的一类，也是最容易被单独使用的一类。

**熟练。**又快又稳，错误率低，不需要停顿。

**记住。**隔一段时间还说得出、写得出。

**完成。**环节走完了：讲了、练了、评了、总结了，作业交齐了。

**作品像样。**有一件拿得出手的产物——海报、模型、报告、汇报。

这五类各自都有道理，谁也不能取消。问题在同一件事上：

**它们全都可以在真正的学习没有发生时出现，而且往往出现得更漂亮。**

一个孩子可以答案全对，而他的正确规则和一条错误规则在这些题上恰好产出同一个数（这本书第七章有一份实测：在一批带误解标注的题库里，同一个知识点上任取两个持不同误解的孩子、任取一道题，**九成以上的情况下这道题看不出他们有什么不同**）。

一个孩子可以极其熟练，而熟练本身正在把这个方法推到它不该去的地方（第六章）。

一个孩子可以记得很牢，记住的却是一句被封死的口诀（第一章）。

一节课可以环节完整、按时收束，而课上真正出现的那个分歧被“课后再讨论”消化掉了（第四章）。

一件作品可以很像样，而它只在作者站在旁边解释时才成立（第四章）。

**五类凭据没有一类是假的。它们全是真的——只是它们证明的那件事，比我们以为的小。**

## 五、地基：以未来为存在条件的东西，当堂验收不了

上一本书第一编界定了三种“真正的学习”到底是什么样的东西。三个名字很硬，我把它们各用一句话翻译过来：

- **再生形式**——一次经验形成的组织原则，**能不能在换了材料之后重新长出来。**
- **可重构敏感域**——学习者判断“什么算差异”的那条边界，**能不能被持续出现的异常改写。**
- **可持续建整体**——一套知识能不能在不断打补丁之后，**还有能力整个重画一次。**

三个对象是分头写的，用了三组完全不同的学问，落在三个不同的地方。但把它们放在一起，会看到一条共同的存在条件——**而这条条件是全书的地基：**

**三者全都以“尚未发生的经验”为存在条件。**

再生形式要看它能否在**新材料**里重建差异；敏感域要看持续残差能否改写**后来的**判断边界；可持续建整体要看重复故障能否触发**下一次**架构重画。

**三者没有一个能在学习结束的那一刻被完整验收**——不是因为测量技术不够好，而是因为它们的存在方式就是“参加以后的事”。

由此得到那条地基性的后果：

**凡以“参加未来经验”为存在条件的东西，就不可能在事情结束的那一刻被完整验收。任何在当堂完成的验收，测到的都只能是它的替身。**

请把这句话记住，因为你在这本书里遇到的每一个“分界落在后面”，都出自它：

第一章的分界在**换掉承重条件之后**；第二章在**撤掉支架之后**；第三章在**材料被替换之后**；第五章在**走偏之后**；第六章在**遇到近邻之后**；第七章在**没有提示的迁移任务上**；第八章在**隔一段时间之后**；第九章在**去掉标签的混编题上**。

**八处，没有一处落在当堂。这不是八次巧合，是那条地基决定的。**

## 两个必须说明白的缺口

这条地基带来两个缺口，上一本书写明了，这本书也一样躲不开。

**第一个缺口：全部依赖行为层面的观察。**

前面那三个对象的性质，最后都要落到某种看得见的东西上——他说了什么、写了什么、在新任务上做了什么。**学生内部已经发生、却没有进入公共记录的变化，这套眼光处理不了。**

这不是谦虚，是一条真实的限制。一个孩子可能在心里完成了一次对象重定，只是不善于说出来；这套眼光会把他记成“没发生”。所以这本书里所有的记录表都强调**跨任务、跨时间**——单次观察的误判率高得不能用来给人下判断。

### 第二个缺口：这些东西不会自动跨领域通用。

一个人在代数上形成了那条能被改写的判断边界，**不代表他在几何上自动拥有**，更不代表他在历史证据或审美判断上拥有。上一本书写明了这一点，也写明了它没有解决：“跨领域重建需要什么条件、要多久”，没有答案。

**所以如果有人把这些当成一种可以脱离内容去培养的一般能力——“思维训练课”——那是这套眼光的误读。**它从头到尾长在具体的内容里。这本书之所以要一章一章地把它放进分数单位、三角形的高、移项，正是因为它离开这些东西就不成立。

## 六、九个靶格放在同一张表上

上一本书九章各有一张二维辨别格，各有一个“最难被识别”的格子。把九个格子拆下来放在同一张表上——这是装订成书才有条件做的一件事：

| 章 | 靶格    | 它凭什么通过检查           | 分界在哪里显露        |
|---|-------|--------------------|----------------|
| 一 | 熟练复现  | 正确率高、保持久、相似任务能迁移   | 换材料、撤提示、延迟之后   |
| 二 | 刚性稳态型 | 表现稳定、抗干扰、术语完整、自信   | 加入一个使原规则失效的变化后 |
| 三 | 补丁固化型 | 错误越来越少、笔记越来越完整     | 多个局部故障指向同一假设时  |
| 四 | 封闭推进型 | 目标清楚、讲练衔接、错误当场处理   | 预设的触发条件出现之后    |
| 五 | 预设变式型 | 变式丰富、迁移题精心设计、成绩良好  | 返回的差异由谁真实产生    |
| 六 | 中央整合型 | 教师倾听敏锐、每个回答都进入后续   | 教师退场之后         |
| 七 | 提前结算型 | 目标、活动、反馈、作品、成绩一应俱全 | 延迟迁移、无提示发起     |
| 八 | 代偿通行型 | 正确率、完成度、流畅性、满意度俱佳  | 外部通路撤除之后       |
| 九 | 差异泄压型 | 纠错迅速、共识清楚、表达提炼     | 新材料中能否主动发现同类张力 |

## 七、一条事后才显露的规律

九章是分头写的，九张表也是分头做的。并排放在一起之后，**有一件写作时并未预设的事显露出来：**

**九个靶格没有一个是明显的坏课堂。**

九格全都在它被评价的那个维度上合格，多数还优于对照。熟练复现比接触性变化分数高；补丁固化比零件堆积错误少；预设变式型比即时清零型设计精良；中央整合型比教师串行型更被称赞；代偿通行型比暴露停滞型完成度高。

**九次比较，靶格全部胜出。**

于是有了那条总判断：

**课堂真正的失败形态，恰好长得像成功。明显的失败有自我报警能力，制度会自动处理它；这九种形态没有报警机制——因为每一种都在被评价的维度上是合格的。**

它有一个可以立刻使用的推论：

**凡是只能在当堂观察到的指标，都无法区分这九格与它们各自的健康对照格。**

这一条直接冲撞现行的课堂评价惯例。听课评课、教学比赛、随堂检测、平台的即时数据，**全部在当堂完成**。它们可以有效识别那种明显的失败，却在原理上识别不了这九格。

**这就是你手上这本书为什么长这个样子。**全书那张格子的左下角——那个“死的正确”——就是上面这九格在数学课堂里的样子。它之所以危险，不是因为它差，恰恰是因为它在所有当堂能看见的维度上都很好。

## 八、不是一个，也不是九个，而是三层三个

九个靶格共享同一个生成条件——一个位置只要能被当堂结清，它就会被当堂结清。那么它们是不是其实是同一个东西？

答案是：**不是一个，也不是九个，而是三层三个。**

**第一层：学习者层。**（熟练复现、刚性稳态、补丁固化）

主语是学生：他能不能跨材料重建、能不能重画判断边界、能不能停止加补丁。**撤掉课堂，这三格仍然成立**——一个自学者、一位工程师、一个组织，同样可能停在这三格里。

**第二层：课堂层。**（封闭推进、预设变式、中央整合）

主语是课堂这个多人系统：结论有没有重开的入口、返回的差异由谁产生、下一轮由谁先动。**撤掉多人结构，这三格不成立**——一个人自学，不存在“下一步由谁先动”的问题。

**第三层：制度层。**（提前结算、代偿通行、差异泄压）

主语是制度与情境：为什么必须结清、为什么必须恢复通行、为什么必须恢复平稳。**它们不描述状态，描述的是把状态往回拉的那股力。**

三层不能互相还原。判据是——**改变一层，能不能单独改善另一层？**

- **只训练学生，不改课堂运行：**第一层的东西没有生长的位置。
- **只改课堂运行，不动那三条力：**新的做法会被迅速改造成它的表演形态。
- **只谈那三条力，不给对象与运行的判据：**批评无法落地，也无法被判错。

这三条不是修辞，是可以做的实验。

**这就是这本书三编的分工。**第一编说的是孩子身上的事（第一层），第二编说的是一堂课怎么运行（第二层），第三编说的是那三股把课堂往回拉的力（第三层）。

读到第三编你会发现语气变了——前两编在说“你可以这样做”，第三编在说“为什么你这样做了也没用”。**那不是作者泄气，是层次换了。**

## 九、三条力：三种关闭方式

第三层那三条力，做的是同一件事——**让一件事不再构成问题**。区别在于关闭的方式：

**第一种，账上关闭。**那一栏必须在结果出来之前填掉，于是“能不能结算”成了前置的选择器。**课时有结尾，单元要验收，成绩要出，责任要落。**学习的后果显露得很慢，而账要现在就结。于是能被现在结清的东西优先被生产出来。

**第二种，绕过关闭。**障碍没有被处理，只是任务从旁边过去了。帮助、提示、模板、工具，在学习者完成障碍定位之前就恢复了任务的输出——**原来那个障碍因此失去了结构上的身份**。而且帮助越成功，下一轮外部系统来得越早。

**第三种，就地降压关闭。**差异被允许出现，然后被转换成不会继续改变结构的形式：先当成一个局部错误处理掉，再变成大家都同意的说法，最后固化成标准表达。**课堂迅速恢复平稳，而产生冲突的那条规则、那个边界，一个字没动。**

在这本书里，这三种关闭方式分别长成了第七、八、九三章：正确来得太早（账上关闭）、余量被清零（绕过与结算的合流）、运输单位取代学习单位（把差异压成可治理的标准形式）。

**三章共同的那个观察，比三条机制本身更值得记住：**

这三条力都随着“做得更好”而加强，不是随着“做得更差”而加强。结算的格式越细密，提前结算越普遍；帮助越精准，代偿越隐蔽；总结越清楚，压缩掉的东西越多。

这句话意味着一件很不舒服的事：**改进本身可以是病因。**

## 三条往前的链，三条往回的链

三层之间不是并列，是**生成关系**：制度那一层决定课堂这一层的成本与资格，课堂这一层决定学习者那一层能不能反复发生。

上一本书把这件事写成了六条链。三条往前，三条往回。

**三条往前的链：**

**重开 → 判断边界。**旧结论被新证据合法地重开，学习者被迫重新决定哪些差异算信号——判断的边界因此被改写。**注意这里的关键不是“提高了参与度”，而是“后来的证据有权修改过去的裁决”。**这本书第五章那条“可改判的临时裁决”，就是这条链。

**回流 → 再生。**有价值的未决差异，穿过时间和材料的变化再次进来，旧经验不再被机械复现，而是在新条件里长出不同的组织路径。这本书第三章的“回写轮”和第四章的“故障回写下一轮”，是这条链。

**换手 → 续建。**知识的状态被封装、被继承、被变形、再被公开地写回去——整体不再依赖某一个中心来保持连续。这本书第四章的“终点换手”是这条链。

**三条往回的链：**

**结算 → 封存。**结算压力让课堂优先完成当前的凭证，于是结论保持开放的时间被缩短，未决差异的容量被压小，重开变成一道走过场的程序。典型结果是“学过、说过、做过、得分过”替代了真正的生成。

**代偿 → 夺取。**帮助在学习者完成障碍定位之前就恢复了任务的通行，外部系统由此获得越来越早的首动权；接手权退化成中央分发任务，判断边界也失去了被残差改写的机会。

**泄压 → 压缩。**差异被局部化、被协调、被标准化，于是它无法跨人、跨时间保持自己的身份；未决差异和接手权同时失去载荷，而知识整体靠不断加补丁维持表面的完整。

**三条往回的链彼此强化。**结算压力要求课堂按时关闭，关闭需要帮助系统快速恢复通行，而通行又需要把难以协调的差异压缩成可处理的局部问题。

**最后形成的，是一个非常稳定的高绩效外观：活动丰富、任务完成、错误减少、记录完整——而学习结构很少改变。**

这段话如果你觉得眼熟，是对的。这本书第三编那三章，讲的就是这三条往回的链在数学课堂里的样子。

### 一个提醒：不要给课堂贴整体标签

上一本书把这九个概念处理成**状态变量**，而不是人格标签。这个处理值得单独说一句，因为它决定了这本书该怎么用。

同一个班级可以在某个数学单元里保留了很多重开的空间，却在另一门课上一点也没有；同一位老师可以在讨论阶段把发起权交出去，在总结阶段又全部收回；同一个人工智能工具可以在一个任务里制造代偿，在另一个任务里通过分级提示保护学生的首动权。

**九个状态说的是“这个课堂事件此刻处在什么组织状态”，不是“这个班好不好”“这个老师行不行”。**

还有一条记账要求，对读教育研究的人很有用：**任何声称“某教学法有效”的主张，都必须同时报出当时那三条力处在什么水平。**一次重开在低结算压力下可能重写课程，在高结算压力下只是一份“学生有质疑精神”的新凭证。**不报环境而只报做法的效果，在这套口径里不构成证据。**

## 十、三对正面相抗：它们争的是同一个位置

把第二层的三条做法和第三层的三条力并起来看，会发现它们不是“好做法”与“坏环境”的并列，而是**三对争夺同一个位置的力**。每一对争的都是同一件事：**某个位置，是现在关掉，还是留到后面。**

| 争夺的位置    | 课堂这一层要的        | 制度那一层要的    | 争的是什么        |
|----------|----------------|------------|--------------|
| 已经形成的结论  | 保留重开的条件        | 现在就结清      | 现在关闭，还是留着能重开 |
| 学习者遇到的障碍 | 由携带差异的人接手先动    | 外部通路尽快恢复通行 | 下一步由谁先动      |
| 尚未消化的差异  | 少量差异挂着，以后改形再进来 | 就地降压，恢复平稳  | 现在压掉，还是留到后面  |

**三对争的都是同一样稀缺物：课堂里“尚未确定”的那段时间。**

这也解释了为什么这本书里每一条做法都带着限额——为什么第三章说“不是每节课都要改写旧知识”，为什么第五章说“一节课值得走完五步的只有一两道题”，为什么第六章说“一道远的正例加一道近的负例，胜过二十道同调的练习”。

**不是因为作者谨慎。是因为这三条做法争的是同一段时间，而那段时间总量有限。谁要得太多，另外两个就没有了。**

## 十一、这套眼光和最近的五种说法差在哪儿

上一本书的两百四十二条参考文献里，只有五条被三章以上引用。它们是：生产性失败、为未来学习做准备、适应性专长、学习与表现的区分、形成性评价。

**这五家共同占据的那片区域，正是“失败长得像成功”这条判断落下的地方。**所以必须说清楚差在哪儿——否则这套眼光只是一次重新包装。

**与“学习与表现的区分”。**这是最近的一家。它指出当堂表现不是学习的可靠指标：制造流畅感的条件常常损害长期保持与迁移。

**分界在于两者说的不是同一件事。**那一家说的是**测量**：表现与学习之间的相关可以为零甚至为负，所以不能用前者代替后者。这本书说的是**生产**：这九种形态不是被误测的学习，而是“必须在当堂可结清”这个条件**生产出来**的东西——它们在当堂合格，不是因为测量此刻不灵，而是因为**合格正是它们的成立条件**。

由此得到一条那一家不会提出的预测：**误差有系统方向**。如果表现只是不可靠的指标，那么“当堂高分+延迟低分”的搭配应该在各类课堂里大致随机分布；而这套眼光预测的是——**在当堂评价维度上得分最高的那一格，恰好是最难识别的失败形态**。前面那九次比较里靶格全部胜出，不是九次巧合。

**与“生产性失败”。**它说的是当堂看起来更差的过程可能带来更好的后续学习。它和这套眼光共享“当堂读数会骗人”这个方向，但站在同一张表的**另一格**：它处理的是表现差而学习好，这九格处理的是表现好而学习差。两者不冲突。

真正的接触点在别处：**生产性失败是一份处方，这套眼光是一份失败形态学**。处方被采纳之后，失败可以换一副样子继续存在——“预设变式型”正是探索环节被提前编排好之后的形态。

**与“为未来学习做准备”。**那一家测的是下一次学习的速度，这套眼光问的是未来经验的入口本身有没有被改写。还有一条更要紧的：那一家改造的是评价方法（给资源、看后续学习）。而第三层那条力预测——**任何一种可在有限时间内完成、又被用作正式依据的评价，都会被结算化**。一旦这类任务成了正式依据，课堂就会开始为它做准备，而准备本身使它失去检验力。

**与“适应性专长”。**那一家分的是人的两种专长：常规专长在熟悉情境中高效，适应性专长能在新情境中重构。这套眼光分的是**课堂把哪一种生产出来**，并且把“适应性”拆成了三件可以分别失败的事。

**与“形成性评价”。**这是最容易被误认为同一件事的一家。**分界在于：形成性评价改变的是信息的流向**——把结果反馈回教与学；而这套眼光要的是**程序上的资格**——预先写下的触发条件一旦出现，旧的裁决必须被重开，而且重开的结果必须回写结构。

**一堂课可以有充分的形成性评价，而完全不重开任何旧结论**：反馈全部用于下一道题，没有一条用于改写上一次的判断。

**五家合起来。**这套眼光最危险的还原路径，不是被其中任何一家单独吸收，而是被读成它们的合成。硬区别只有一条——**方向**：

**那五家都是改进的处方，说的是“这样做会更好”。这套眼光是失败的生成学，说的是“不这样做，失败会长成哪九种样子，而且长得像成功”。**

**处方可以被采纳，而失败依然存在。**这正是这套眼光能被判错、而那种合成读法不能被判错的地方。

## 十二、它看不见的那一格

任何一副眼光都有它照不到的地方，这一副的那一处，我直接写出来。

那九张表全部以“课堂”或“学习者”为单位，**没有一张处理学生之间的分布**。于是有一种失败形态落在九格之外：

**课堂整体的读数都不差，但收益在学生之间高度不均——而且这种不均，恰恰由前面那些机制的正常运行产生。**

它同样难以被识别（班级均值很漂亮），却不落在任何一格里。上一本书给它起了个名字叫**分配性失败**，给了它一个读数，但那只是一个读数，不是一层理论。

我把它写在这里，是因为它对数学课堂特别要紧。这本书里几乎每一条做法——留三秒、让孩子先说、造一个近邻、把差异写进下一轮——**在实施的时候都天然偏向那些已经能开口的孩子**。一个从来不举手的孩子，你给全班留三秒，他还是不会开口。

**这一格，这本书解决不了。但你在用书里任何一条做法的时候，都该把它挂在心上：这一次，是谁得到了那三秒？**

### 还有一条规矩：这些东西不能用来做什么

上一本书在卷首写了一条声明，我把它原样搬过来，因为它对这本书同样有效，而且更要紧——这本书的九张操作卡比那本书的指标更容易被拿去用：

**这套读数只适用于研究、课程设计、共同备课与制度诊断，不得直接用于教师、学生或学校的高风险排名与奖惩。**

理由不是出于谨慎，是出于这套眼光自己的判断——**指标一旦成为高风险目标，最先被生产出来的一定是该指标的表演形态。**

请把这句话和第八章那条对照着读：三条力都随着“做得更好”而加强。如果有人把“余量存活时长”“差异回写率”“拆箱重构率”做成一张考核表，那么下个学期你会看到的，是被精心制造出来的余量、被精心安排的差异、以及被专门设计来展示“拆箱”的公开课。

**这本书会亲手造出它自己要对付的那第十格。**

还有一条限额上的规矩。上一本书的三条课堂机制，各自都拒绝了“越多越好”，而且给的理由是同一个：**它们争的是同一段稀缺的时间**。重开太频繁，课堂无法形成任何稳定的结论；未决差异挂得太多，容量会崩；换手太密，接手者来不及形成自己的判断。

三条都是**倒U形**：太少不行，太多也不行。

这本书里每一条做法后面那句“一节课挑一处就够”“一个单元真正值得改写的只有一两个节点”“两个候选就够”，都是这条倒U形的具体值。**它们不是保守，是这套理论自己推出来的限额。**

## 十三、九个术语，一句话一个

这是上一本书的九个核心概念。你在这本书里不会再遇到这些名字（这本书用的是它们在数学里的样子），但如果你要回去读那一本，这张表就是钥匙。

| 层   | 术语   | 一句话                    |
|-----|------|------------------------|
| 学习者 | 再生形式 | 换了材料之后，那个组织方式还能不能重新长出来 |

|     |        |                         |
|-----|--------|-------------------------|
| 学习者 | 可重构敏感域 | "什么算差异"这条边界，能不能被持续的异常改写 |
| 学习者 | 可续建整体  | 补丁堆到一定程度，能不能整个重画一次      |
| 课堂  | 可复议推进  | 结论带着自己的重开条件往前走          |
| 课堂  | 限额变奏回流 | 少量未决的差异获得未来资格，改形之后再进来   |
| 课堂  | 携差换手   | 下一步由携带着差异的接手者先动         |
| 制度  | 结算先行   | 账必须在后果显露之前结清            |
| 制度  | 代偿通行   | 障碍被绕过，任务恢复了通行           |
| 制度  | 差异泄压   | 张力被释放，结构没有改变            |

## 十四、两本书的对照表

最后这张表，是这一章真正的用处：上一本的九格，在这本书里长成了什么样子。

| 层   | 《课堂的智慧》的靶格 | 本书对应的那一格 | 在第几章 |
|-----|------------|----------|------|
| 学习者 | 熟练复现       | 封死的正确    | 第一章  |
| 学习者 | 刚性稳态型      | 过界的熟练    | 第六章  |
| 学习者 | 补丁固化型      | 标本式的正确   | 第三章  |
| 课堂  | 封闭推进型      | 导航式的正确   | 第五章  |
| 课堂  | 预设变式型      | 蒸发的正确    | 第四章  |
| 课堂  | 中央整合型      | 借来的正确    | 第二章  |
| 制度  | 提前结算型      | 安静的正确    | 第七章  |
| 制度  | 代偿通行型      | 高效空心     | 第八章  |
| 制度  | 差异泄压型      | 封箱的正确    | 第九章  |

请注意这张表不是一一对应的翻译。左边是课堂的一般形态，右边是它在数学这门学科里的具体长相；有几处的对应是斜着的——比如"刚性稳态"在一般课堂里是"规则不肯改"，落到数学里就成了"一个太顺手的方法往外扩张"。

两本书的关系可以用一句话说完：上一本回答"课堂为什么会这样"，这一本回答"那么这道题该怎么问"。

### 三处对应得最紧的地方

九对里有三对贴得特别近，值得单独说，因为它们是两本书之间真正的接榫。

**第一处：中央整合型 ↔ 借来的正确。**

上一本书里，"中央整合型"说的是一位倾听极其敏锐的老师——每一个学生的回答他都能接住、都能编进后面的推进里。这样的课堂在任何一次评课里都是范例。它的问题只有在教师退场之后才显露：**每一条线都从他那里过，于是没有一条线属于学生。**

落到数学里，它就成了这本书第二章的样子——**答案是真的，撑着它的柱子却不在孩子这一边。**第二章那六根柱子（题面的词、措辞、画出来的线、一个符号动作、章节标题、老师的表情），最后一根正是"中央整合"在一道题上的具体长相。

### **第二处：预设变式型 ↔ 蒸发的正确。**

"预设变式型"说的是变式设计得极其精良的课：题目层层递进，迁移题精心安排，成绩也好。它的问题在于——**返回课堂的那些差异，全是老师提前编排好的，没有一个是学生真正产生的。**

落到数学里就是第四章：讨论很热烈，最后老师总结出正确的结论，而**第二天的练习册跟昨天那场争论毫无关系。**那个数字我在第四章给了：本意在于"建立联系"的题目，最后真的建立起联系的比例，在有的国家不到百分之一。

### **第三处：提前结算型 ↔ 安静的正确。**

"提前结算型"是那种目标、活动、反馈、作品、成绩一应俱全的课。它是九格里最难指摘的一格，因为它什么都没做错。

落到数学里是第七章：**全班都对了，于是这个班此刻只有一份判断。**而第七章那次真算给了它一个数字——在一批带误解标注的题库里，一道题平均只能让四成五的已知误解显形。

**这三处说明了一件事：两本书不是同一本书的两个版本。**上一本给的是形态学——这九种失败长什么样、怎么区分、在哪儿显露；这一本给的是**这些形态在数学这门具体学科里的落点，以及一位老师明天能对它做什么。**

形态学不告诉你明天该问哪一句话。这本书的九张操作卡才是。

## **十五、读完这一章之后**

如果这一章你只带走三样东西，我建议是这三样：

**第一，那条地基。**凡是以"参加未来经验"为存在条件的东西，就不可能在事情结束的那一刻被完整验收——当堂测到的，永远只是它的替身。这一条解释了这本书里所有"分界落在后面"的判断。

**第二，那条总判断。**课堂真正的失败形态，恰好长得像成功。明显的失败会自己报警，制度会处理掉它；这九种不报警，因为每一种都在被检查的维度上合格。

**第三，那条分层。**学习者、课堂、制度，三层不能互相还原。**你在哪一层，能做的事就不一样。**一位老师能在第二层做很多事，第三层的多数事情不在他的权限里——这本书结论那一章之所以要分开给三种人三句话，就是因为这条分层。

至于九个术语，你可以忘掉。它们是上一本书的语言。**从下一章开始，这本书只用一张格子和九个更好记的名字。**

## **十六、这一章的一句话**

明显的失败会自己报警，制度会替你处理掉它。这本书对付的是另外九种——它们不报警，因为它们在每一个被检查的维度上都合格。

下一章开始，我们回到具体的题目上。第一编三章问同一件事的三个侧面：**没走的那条路还在不在、是谁在替他撑着、旧知识有没有被改写。**

# 第一编 什么才算真的学会

第一至三章

# 第一章·没走的那条路还在不在

第一编 什么才算真的学会·之一

## 一、两个都得满分的孩子

下面这个场面是我构想的，但每一处细节都来自真实课堂里常见的情形。

四年级，三角形面积。

题目给了一个三角形，底边 8 厘米，高 5 厘米，求面积。

小舟看了一眼，写下  $8 \times 5 \div 2 = 20$ ，单位写好，打上句号。用时十一秒。

小原也做对了，用时二十几秒。中间他停了一下，用铅笔在图上把那条高描了一遍。

两份作业摆在一起，一模一样：答案对，单位对，字迹清楚。老师各画一个勾。从这一刻的证据看，两个孩子处在同一个位置。

现在我们换一道题。

还是那个三角形，但是把它**转了一下**，让原来那条底边斜着朝上。

小舟愣了两秒，开始找“哪条是底”。他最后选了图里最下面那条线——不是因为它是底，是因为它在下面。

小原也停了两秒，然后说：“我得先决定拿哪条边当底，然后再从对着它的那个顶点画下来。”

再换。这次不告诉高，只告诉你：这个三角形的顶点，在一条和底边平行的直线上。

小舟卡住了。他说：“没有高，做不了。”

小原也没有马上做出来。但他做了一件事：他画了一条虚线，从顶点垂直落到底边所在的直线上，然后说：“高应该是这两条平行线之间的距离，顶点在这条线上随便动，面积都不会变。”

这一次，两份作业不一样了。

但我想请你注意一件更要紧的事：**这个差别，在第一道题的时候就已经存在了。**只是那张满分的卷子，没有地方可以把它记下来。

## 二、这一章要还的债

**做对一道题，可能发生两件完全不同的事：一件是路走通了；另一件是路走通了，而其他没走的路，全被封死了。**

后面整章，我都在还这句话的债——它到底指什么，怎么看得出来，以及为什么它比“理解得深不深”更值得问。

## 三、三门学问，三个不一样的答案

我要从三门离数学很远的学问进来。它们各自都认真回答过一个问题：**什么才算真的存在？**

而且它们的答案互相不服。

**植物学：一棵树，不等于它此刻伸出来的枝条**

**先说清楚植物学在这里研究什么。**

我们通常以为，一棵树往哪儿长是“天生的”。植物学研究的恰恰是另一件事：**一根枝条长不长出来，是怎样被决定的。**

关键的部件叫**腋芽**——就是叶子和茎交界的那个小夹角里，藏着一个针尖大的芽点。它是一根潜在的枝条。一株植物身上有几十上百个腋芽，绝大多数一辈子都不会长出来。

为什么不长？因为顶端在压着它。植物学把这叫**顶端优势**：最顶上那个芽会持续向下发出信号，抑制下方的腋芽萌发。这不是玄学，是可以做实验的——**把顶芽剪掉，过几天，下面原本睡着的腋芽就开始动了。**

这件事我们其实都见过。家里的月季或者绿萝，掐掉顶尖之后，旁边会冒出两两根新枝，长得更密。园艺上管这叫“打顶”，其背后就是顶端优势被解除。

所以对我们有用的是这三个词，各用一句话记住：

- **腋芽**：藏着的、还没长出来的枝条；
- **休眠**：它没长，不等于它死了；
- **顶端优势**：它没长，是因为有别的东西正在压着它。

第三句最要紧。“**没长出来**”和“**长不出来**”是两件完全不同的事——而在一张卷子上，这两件事看起来一模一样。

我们看一棵树，最自然的理解是：它就是眼前这些茎、叶、花、果的总和。

植物学不这么看。

很多开花植物的叶腋处有一种组织，能长出腋芽。一个腋芽可能很快长成侧枝，也可能长期停在休眠状态。它长不长，不只取决于芽本身，还取决于顶端优势、糖分供应、光照的性质、温度、养分、激素，以及其他枝条正在干什么。

更要紧的是研究者发现：芽从休眠转向活跃，**不是“里面早就装好了一根小枝，外界只负责按开关”这么简单。**芽的生长潜力，是在整株的信号和局部条件不断整合的过程中获得的。剪掉顶芽、改变光照、调整养分，都可能重新分配枝条的机会。

所以一棵树的“形”至少有三层：已经长出来的枝条；生长过程中被压制、延缓、舍弃的方向；以及条件一变仍然能被激活的分枝能力。

只看第一层，我们会以为这棵树已经把自己完全写在外表上了。

**但植物学同时也不给我浪漫的空间。**有人读到这里会说：所以教育就是相信每个孩子都有无限潜能。植物学恰恰不支持“无限”两个字。芽的数量有限，位置有限，养分有限；有些芽已经失去活性，有些即使萌发也会争养分，让整株变脆弱。植物不是把所有可能都实现，而是在特定环境里**保留、抑制、释放和修剪**可能。

这一条约束，后面会砍掉我一句很想说的话。

## 考古学：为什么他们更关心废墟，而不是成品

先说清楚考古学在这里研究什么。

一提考古，容易想到墓葬和青铜器。但这里要用的是考古学里一个很朴素的分支：**研究一件东西是怎么被做出来的。**

它有个专门的名字，叫**操作链**（法文 chaîne opératoire，直译就是“一连串动作”）。做法是：不只看成品，还要把制作过程中掉下来的所有东西都捡起来——原料的边角、敲下来的碎片、做坏了的半成品、断掉的部分——然后从这些东西倒推出当时那个人的动作顺序。

打制石器最能说明问题。一块石头要做成石斧，敲下来的碎片（叫**剥片**）有几十上百块。这些碎片散落在地上，看起来就是垃圾。但研究者发现：

- 不同阶段敲下来的碎片，**形状和厚度不一样**——所以从碎片能读出他做到哪一步了；
- 新手和熟手敲出来的失误痕迹**不一样**——所以从碎片能读出他熟不熟练；
- 碎片在地上的**分布位置**也有讲究——有的遗址里，碎片的聚集方式被用来讨论"这里是不是有人在教另一个人做"。

换句话说：**成品告诉你他做成了什么，碎片才告诉你他是怎么做的。**

这对我们的用处直接得让人不舒服。课堂上，草稿纸上那些涂改、试算、画错的辅助线、划掉的设元——**那就是碎片**。而标准答案是一件被打磨干净的成品，它把所有岔路都清掉了。

设想一万年有人打制了一件石器。

今天，成品摆在博物馆里，对称、锋利、尺寸合适。我们很容易赞叹这个人的手艺，却很难仅凭这件成品知道：原料从哪儿选的，第一锤落在哪里，中途改过几次方向，哪些剥片是有意的，哪些是失手，这个人有没有被熟练的人指点过。

考古学为此发展出对"操作链"的分析：不只看最后那件器物，还从原料、石核、剥片、废屑、半成品、断裂的方式和它们散落的位置，重建一件东西从选料、制作、使用到废弃的一生。石器研究里尤其明显——地上那些碎片不是成品旁边没意义的垃圾。某些剥片能反映制作阶段，某些错误痕迹能暴露操作者的熟练程度，某些空间聚集甚至被用来讨论学徒活动和示范教学。

于是考古学给了一个和学校评价完全相反的判断：

**成品越完美，过程有时越难被看见；真正暴露手艺结构的，往往是那些没有进入成品的东西。**

课堂恰好倒过来。孩子草稿纸上的涂改、试算、画错的辅助线、被划掉的设元，常常被当成"脏乱"。而标准答案会把过程洗成一条没有岔路的路：因为 A 所以 B，由 B 得到 C，最后求得 D。它给后来的人一种错觉，好像解题的人从一开始就知道终点在哪里，从来没在错误的对象上停留过。

**但考古学同样限制我。**它比谁都清楚：一样的痕迹可能来自不一样的过程。一堆碎片不会自动说明什么。这一条，等一下会砍掉我另一句想说的话。

## 绘画：被覆盖掉的线条，还算是作品的一部分

**先说清楚这里用的是绘画研究的哪一块。**

不是画得好不好看，是一门很技术的活儿：**用仪器看一幅画的底下有什么。**

油画是一层一层画上去的。画家改了主意——把一条手臂挪下去三寸、把头转个方向、把一根柱子整个盖掉——被盖住的那部分并没有消失，它还在下面。意大利语管这种"改主意的痕迹"叫 **pentimento**（字面意思就是"后悔"），复数是 **pentimenti**。

**怎么看见它？**两个办法，都不用破坏画：

- **红外反射成像**——红外线能穿透上面的颜料层，照出底下的炭笔起稿线；
- **X 光**——含铅的白色颜料挡 X 光，所以能照出被覆盖的形体。

各大美术馆的技术报告里，这类记录多得是：一幅完成了几百年的画，底下常常压着好几个被放弃的构图。

于是美术史上就有了一个不太好回答的问题：**这些看不见的线条，算不算这幅作品的一部分？**

从观看的角度，它们不在画面上，观众看不见。但从形式的角度——最终那个构图之所以是这样，正因为它**拒绝过那些**。画家把手臂往下挪了三寸，最终的姿态才有了那种张力。

一句话记住：**被拒绝的形式，参与了最终形式的形成。**

在西方绘画研究里有个词叫 *pentimento*，字面意思是“悔改”，指画家在作画过程中改掉的地方——挪过位置的手臂，换过方向的头，画到一半被盖住的柱子。

有意思的是，这些改动往往并没有消失。颜料层会随时间变得半透明，红外反射成像也能把底下的层次照出来。国家美术馆的技术报告里有大量这样的记录：一幅完成了几百年的画，底下压着好几个被放弃的构图。

问题来了：**这些被覆盖的线条，算不算这幅作品的一部分？**

美术史给的答案很微妙。从观看的角度，它们不在画面上，观众看不见；但从形式的角度，最终那个构图之所以是这样，正是因为它**拒绝过那些**。画家把手臂往下挪了三寸，最终的姿态才有了那种张力。被拒绝的形式参与了选择。

**但绘画也给我一条限制**，而且是最不客气的一条：改动本身不高贵。反复修改可能来自技法不足、判断摇摆、迎合别人，或者单纯拖延。一个人画了二十稿，不一定比一稿画完的人更懂形式。

## 它们打架的地方

在往下走之前，先把三门学问的词收在一张表里。

| 这个词               | 一句话解释          | 在这一章里对应什么    |
|-------------------|----------------|--------------|
| 腋芽                | 藏在叶腋里、还没长出来的枝条 | 没走的那条解法      |
| 休眠                | 没长，但还活着，条件变了能长 | 还能重新接回来      |
| 顶端优势              | 被别的东西压着，所以不长   | 主方法太顺，别的路被封住 |
| 操作链               | 从碎屑倒推制作过程的方法   | 从草稿倒推思考过程    |
| 剥片                | 制作中敲下来的碎片      | 涂改、试算、画错的线   |
| <i>pentimento</i> | 被颜料盖住的、改过的笔迹   | 被放弃却参与了选择的想法 |

现在把三家的答案摆在一起：

- 植物学说：**没长出来的东西也算存在**，因为它保有再长出来的能力；
- 考古学说：**没进入成品的东西才最说明问题**，但你得从留下的痕迹去重建，而痕迹会骗人；
- 绘画说：**被覆盖的选择塑造了最终形式**，可是改得多不等于想得深。

三家不但答案不同，连“凭什么算数”的标准都不同：一个看能力，一个看痕迹，一个看它有没有参与选择。

但三家有一个地方是一致的——而这个一致的地方，恰恰是我要动的地方。

三家都在追问“什么才算真的存在”。而我们的课堂，从来不问这个问题。**课堂的账本上只有一栏：已经实现出来的东西。**写出来的答案，走完的步骤，用得上的方法。

一个孩子心里那条“我本来想这么做，但我知道为什么这条路现在走不通”的路，在这本账上没有位置。它不是被记成零分，是**根本没有那一栏**。

## 四、给它一个名字，和一张格子

我把这类东西叫作“余枝”——不是多余的枝条，是当前这道题做完之后，没有进主干、却仍然能重新长回来的那些关系。

一条余枝要成立，得同时满足三个条件。少一个都不算。

**第一，它没有在这次的答案里实现出来。**

已经写出来的第二种、第三种解法，可以当作余枝存在的证据，但它本身不是余枝。余枝首先是一种**还没有被这次结果耗尽的可能**。

**第二，它和走通的那条路之间，有说得清的排除关系。**

孩子知道自己为什么没走那条路，或者知道那条路要成立需要什么条件。**没有排除理由的“我还可以试试别的方法”，只是模糊的愿望。**这一条是全书最容易被跳过、也最关键的一条。

**第三，关键条件一变，它能重新接回来。**

什么叫关键条件？是那种一旦变了，就逼着你重新决定“研究的是什么、以什么为一份、哪两个量之间建立关系”的条件。把 24 改成 36 不算变；把“每份多少”改成“多少为一份”、把等腰的条件撤掉、把每个面出现机会一样改成不一样、把函数图像换成变化表——这些才可能逼出重新组织。

所以，一句话：

**余枝，是这次答案里没有实现、却因为和主路径之间有说得清的排除关系，而能在关键条件变了以后重新长出对象、单位和关系的那种可能。**

回到序言里那张格子。横轴是这次做对了没有，竖轴换成一句更具体的话：**关键条件变了以后，还长得出来吗？**

|       | 这次做对了                           | 这次没做对                         |
|-------|---------------------------------|-------------------------------|
| 还长得出来 | 活的正确：题做对了，条件一变还能重建              | 正在发生：答案不完整甚至错了，但已经有了说得清的岔路和排除 |
| 长不出来  | 封死的正确：答案、步骤、分数都很好，可原来那条路之外什么也没有 | 普通的不会                         |

传统评价最容易认出来的，是右下和左上：一个明显不会，一个明显会。

**真正危险的是“封死的正确”——它就贴在“活的正确”旁边，两者在卷面上一模一样。**

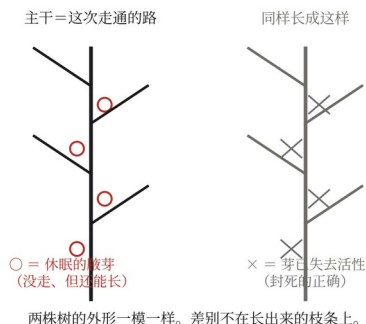


图2 两株外形相同的树。差别不在长出来的枝条上。

小舟不是不会。小舟很会。他十一秒就算完了，比小原快一倍。他的问题在于：在他那次正确里面，“底和高是我为了建立面积关系而选的一对标准”这条关系，已经不存在了。留下的是“图里下面那条线×图里那条竖线÷2”。

而右上角那格——我叫它“正在发生”——也常被误读。孩子可能没做完，却已经发现原来的单位不管用了；可能算错了，却看出两种解法共享同一个关系；可能提了一个还没验证的猜想，却能说清楚原方法到哪儿就走不下去了。这些都不能直接算作正确，**但也不该和毫无结构的失败混为一谈。**

教育真正的难处，不是把所有孩子直接送到左上角，而是：让右上角的发生有机会定型，同时防止左上角的成功把枝子全封死。

## 五、六个场面，六种不同的“重新长出来”

有了这张格子，我们需要一个不靠“深不深”“真不真”“灵不灵活”来判断的问法。

因为这些词都很动听，却没办法进课堂记录。一个只能凭老师印象判断的标准，很快会变成新的标签。

我用的问法是这一句，没有一个形容词：

**把题目里的一个关键条件、一种表示方式或者一个单位换掉之后，孩子的记录里有没有出现：新的对象、新的单位、新的关系，以及原来那条路的失效点？**

四项不要求每次都出现。不同的数学内容，重新长出来的地方不一样。下面六个场面，正好长在六个不同的位置。（以下场景均为构想，题目本身是真实的常见题型。）

### 场面一：整数乘法 —— 长出来的是“对象”

原题： $24 \times 15$ 。

一个孩子用竖式算对了。把15说成“3个5”，他仍然坚持列竖式，看不出 $24 \times 15$ 和 $24 \times 5 \times 3$ 之间有什么关系。

另一个孩子也列竖式，但在解释的时候把15拆成 $10+5$ 。接着把题目改成“24个15和15个24是不是一样多”，他通过把点阵图转九十度重新理解了交换律。

这里重新长出来的是**对象**：乘法不再只是一条运算指令，它成了等组、阵列和面积。

## 场面二：分数加法 —— 长出来的是"单位"

原题： $\frac{1}{3} + \frac{1}{6}$ 。孩子会通分，算对了。

现在把问题改成：一条绳子的  $\frac{1}{3}$ ，和另一条**长度不一样**的绳子的  $\frac{1}{6}$ ，能不能直接相加？

孩子开始意识到：这两个分数的"整体"不是同一个。

这里重新长出来的是**单位**。分母一样还不够——先得确定谁是那个"1"。这是整个小学数学里最容易被符号盖住的一件事。

## 场面三：三角形面积 —— 长出来的是"关系"

就是本章开头那道题。给底和高，孩子代公式；把三角形转过来，把高移到图形外面，再撤掉画出来的高。

孩子能不能通过平行线、等积变形或者补形，重新把面积关系建立起来？

这里重新长出来的是**关系**：底和高不是图上贴着的两个标签，是**为了建立面积关系而选定的一对垂直关系**。

## 场面四：方程 —— 暴露出来的是"失效点"

原题： $3x + 5 = 20$ 。孩子移项，求解，正确。

改成  $3x + 5 = 3x + 5$ 。再改成  $3x + 5 = 3x + 6$ 。

孩子是仍然在找"x 等于几"，还是能把要找的东西换成**解的集合**，认出一个是恒等、一个是矛盾？

这里暴露的是**原来那条路的失效点**：移项没有错，它只是不再产生唯一的数值了。很多孩子在这里说的是"算不出来"，而正确的说法是"所有数都行"或者"没有数行"。这两句话之间的距离，就是这一章讲的全部东西。

## 场面五：概率 —— 长出来的是"权重"

原题：掷一枚均匀的骰子，出现偶数的概率是多少？孩子写  $\frac{3}{6}$ 。

现在把骰子换成一个六个面大小一样、但质心偏了的东西，各面落下的机会并不相等。

孩子还会不会把"六个面"直接当成六个等可能的结果？

这里要重新长出来的是**样本空间和权重**：结果的个数，不自动等于概率的单位。

## 场面六：相遇问题 —— 长出来的是"谁和谁之间的什么"

两个人从两地相向而行。孩子分别算各自的路程。

如果他能把"两人之间的距离在变短"看成**一个变化的对象**，就会长出"速度和"；在追及问题里，把两者之间的间距看成**一个对象**，就会长出"速度差"。

速度和与速度差，不是两个要背的公式，是**对象重新组织之后显露出来的关系**。

只记住"相向用和、同向用差"的孩子，有两条封死的主干。能从"我要研究谁和谁之间的什么变化"重新出发的孩子，有一组还活着的余枝。

## 六个场面合起来说明什么

答案各不相同：乘法重建对象，分数重建单位，面积重建关系，方程暴露边界，概率重建权重，相遇重建“谁和谁之间”。

这个问法没有把数学学习简化成一个统一的动作，而是让我们看见：在不同的内容里，究竟是哪一根关系被换掉了。

## 六、明天就能用的一张卡

要看见余枝，评价里至少得多三类任务。它们不需要额外的课时，可以直接长在你现有的作业和提问里。

### 第一类：换掉一个关键条件（不是换数字）

- 把“每人分得同样多”改成“每组人数不同”；
- 把等腰三角形改成一般三角形；
- 把每种结果机会相等，改成不相等；
- 把只有一个解的方程，改成恒等式或矛盾式；
- 把固定的单位，改成可以变的单位。

### 第二类：问一句失效点

不只问“这道新题怎么做”，还要问：

“原来那个方法，从哪一步开始不能往下走了？”

这一问是整张卡里最省力、也最有效的一问。孩子偶然找到新方法，不等于他保住了旧路的边界。能说出失效点，才说明主干和余枝之间真的有排除关系。

### 第三类：隔一段时间，再来一次

真正的余枝不该只在刚讲完的时候出现。隔两三周，换一种表示方式，触发同一个结构，看孩子能不能重新长出来，而不是背出原话。

一张最简单的记录表（我建议只记事实，不打分）：

| 看什么     | 课堂上的证据长什么样              |
|---------|-------------------------|
| 对象重定了吗  | 他重新说明了“现在研究的是什么”        |
| 单位重定了吗  | 他指出“这里以什么为1”变了          |
| 关系重建了吗  | 他没有沿用原式，重新建立了数量或图形关系    |
| 失效点找到了吗 | 他指出旧方法靠什么成立、在哪一步断了      |
| 枝子回生了吗  | 隔一段时间，曾经被排除的那条路重新成了有效路径 |

### 这张表不能变成分数，更不能用来给孩子分班。

它说的是一次任务里的结构位置，不是一个人的固定品质。同一个孩子可能在分数上枝繁叶茂，在几何上封得很死；今天没长出来，也不代表以后不会。

这里有一条硬规矩，请你替我守住。一个指标只要进了考核，最省事的达标方式一定是制造表面痕迹：草稿更多，解法更多，反思更长。谁要是拿这张表去数“余枝数量”再排个名次，这张表当场就变成它自己要对付的东西——术语更漂亮，表格更完整，课堂上什么也没长出来。

这张表的用法只有一种：**你自己看，用来决定下一句话怎么问。**

## 七、这套眼光的分寸

### 三句我原本想说、但被那三门学问砍掉的话

**第一句：“余枝越多，学习越深。”**这句不成立。

多解法教学有它的价值：比较不同程序，确实能帮孩子看见方法的效率和适用范围。但“多”本身不可靠——材料太多、比较的焦点不清楚的时候，孩子可能只记住表面差异，甚至更少去主动解释。

植物学在这里逼我让步：让所有芽同时萌发的植物会争养分、徒长、变脆弱。数学学习同样需要修剪。一个孩子写出五种公式套法，不如另一个孩子清楚知道：当“总量固定”改成“单位量固定”的时候，原来那个比例关系为什么会反过来。

所以余枝的价值不由数量决定，由能不能回生、有没有排除关系、值不值那个成本共同决定。

**第二句：“把错误都保存下来就行了。”**这句也不成立。

最容易操作的做法是让孩子记错题本。于是错题被剪贴得整整齐齐，原因栏写着“粗心”“概念不清”“下次认真”，改正栏抄一遍标准答案。错误保存下来了，什么关系也没留下。

考古学早就警告过：废屑不会自动说明过程。一条错误要成为余枝的入口，得暴露出一个可辨认的对象、单位或关系上的选择（不是笔误），孩子得说得出来这条路为什么现在被排除，而且条件一变它可能重新成立。

举个正面的例子：解鸡兔同笼时先假设“全是鸡”。这个假设和事实不符，但它不是普通错误，是一次有意识的标准化——先把对象统一，再用腿数的差把对象的差别恢复出来。被排除的“全是鸡”没有作为事实留下来，却作为一种关系操作参与了解法。而把35抄成53，哪怕错满一页，也长不出任何余枝。

**第三句：“只要让孩子经历开放探索，学习就更真实。”**这句最不成立。

绘画告诉我们：改动本身不高贵。有些孩子不断换方法，是因为每一种都没有形成稳定关系；有些孩子坚持第一种方法，是因为他已经看见了它的边界，没必要多做表演。

**没有形成选择、排除和回生关系的开放探索，只留下疲惫和随机。**真正的开放不是每条路都走到底，是知道哪些路为什么关上了，并且保留重新打开它们的条件。

### 四个不适用的地方

**一、基础自动化的阶段，不能被无限打断。**数的事实、基本运算、符号书写需要练到不用想。每一步都要求重新发明，工作记忆会被占满，更高层的关系反而出不来。

**二、时间极紧、以可靠为先的场合。**考试里的基础题，工程里的标准检查，医疗剂量的常规运算——这些时候首先要的是稳、准。这一章讲的是怎么判断学习结构，不是要求每次执行都把枝子展开。

**三、先备知识不够的时候，开放的枝子只是噪声。**没有足够的事实、概念和表示方式，孩子很难形成有约束的替代路径，“自己探索”会退化成瞎试。必要的示范和讲解不能取消。

**四、有些内容的价值就在于高度规范。**证明有严格规则，符号系统需要共同约定。余枝不能成为拒绝规范的借口——真正能回生的路，最后必须接受定义、逻辑和证据的约束。

还有一条，不是关于内容，是关于人：**一次访谈不能诊断一个孩子**。表达能力、紧张程度、和老师的关系，都会影响他说不说得出来。判断必须结合多次任务、书面痕迹和隔一段时间之后的表现。

## 这一章的一句话

**答案不是学习的终点，答案更像一次暂时定型。学习有没有真的发生，要看定型之后，还有没有东西能重新长出来。**

下一章我们换一个角度问同一件事：这个孩子做对了——**那么，是谁在替他撑着？**

## 第二章·这个答案，是谁在替他撑着

第一编 什么才算真的学会·之二

### 一、一道相遇题，两种完全不同的会

下面这个场面是我构想的，题目本身是四五年级最常见的一类。

**甲车每小时行 60 千米，乙车每小时行 50 千米，两车从相距 330 千米的两地同时相向而行，几小时相遇？**

两个孩子都写出： $330 \div (60 + 50) = 3$ （小时）。

问他们为什么这么做。

第一个孩子说："相遇问题就用路程除以速度和。"

第二个孩子说："每过一小时，两车之间的距离一共少 110 千米；330 里面有 3 个 110，所以 3 小时以后距离减到 0。"

两句话都对。第一句甚至更简短，考试的时候更省时间。

现在我把题目改一下：**两车不是同时出发，甲车先开了一小时。**

第一个孩子会问："那还能不能用速度和？"——他在找一条规则，来判断自己那条规则还能不能用。

第二个孩子会先算：甲车已经跑了 60 千米，两车之间还剩 270 千米，然后剩下的过程里每小时仍然减少 110 千米。**他不需要一条新规则，因为他从头到尾研究的的就是"两车之间的距离怎么变"。**

再改一次：两车同向而行，甲车在后面追。

第一个孩子要重新调用一条规则："同向用差。"如果他记得，就做对；不记得，就卡住。

第二个孩子还是那句话：**间距每小时变化多少？**只是这一次，间距是在以每小时 10 千米的速度缩短。

到这里，两个孩子的差别已经很清楚了。但我要说的重点不在"谁理解得更深"——那句话太笼统。

重点是：**第一个孩子那次正确，是靠"相遇"这两个字撑着的。**

题面上出现"相向而行"，他就知道用加；出现"追及"，他就知道用差。**替他完成关系判断的，不是他，是题目的措辞。**这次成功当然是真的成功，只是我们把它整个记在了孩子名下，而没有查清楚：这次到底是谁在承重。

### 二、这一章要还的债

一次成功，不等于一次能力。要判断学习有没有发生，先要查清楚：这个答案，是由孩子的哪一条关系撑起来的，又有哪几分是环境替他撑着的。

### 三、三门学问，三种"凭什么算有能力"

物理学：先问"我怎样改变它"，再问"它是什么"

先说清楚物理学在这里争的是什么。

不是公式，是一个更基本的问题：**一个量，到底属不属于这个物体？**

拿速度举例。我们习惯说“这辆车的速度是每小时 60 公里”，好像速度是贴在车身上的一个标签。但坐在旁边并排同速行驶的另一辆车里的人会说：这辆车没在动。

两个说法都对。因为**速度从来不是车的属性，是车和某个参照系之间的关系**。一句在物理上完整的话必须是：“相对于路面，在这段时间里，用这种方法测得每小时 60 公里。”

有一位数学家诺特在 1918 年把这个思路推到了极处。她证明了一件很漂亮的事：**每一条守恒定律，背后都对应着一种对称性**。

听起来抽象，其实可以这样理解。“对称”在这里的意思是“换个说法它不变”：

- 今天做的实验和明天做，结果一样（**时间平移对称**）→ 对应**能量守恒**；
- 在这里做和挪到十米外做，结果一样（**空间平移对称**）→ 对应**动量守恒**。

诺特的意思是：一个量之所以守恒，**不是因为它自己特别结实，而是因为整个系统允许某一类变换而不改变。守恒是被一整套结构撑着的**。

还有一半物理学更谨慎的态度，常常被我们丢掉：**一条规律有它的适用范围**。胡克定律说弹簧的拉伸和拉力成正比——但那是在弹性限度以内；拉过头，这条关系不但不准，还会彻底失效。牛顿力学在日常速度下极其好用，接近光速就不行了。

所以成熟的物理知识包含三件事，而不是一件：**什么变了；什么在变化中保持；这种保持在什么范围之外不再成立**。数学教学长期只拿走了中间那件。

日常语言里，我们把性质当成物体内部已经有的东西：一块铁有质量，一根弹簧有弹性，一辆车有速度。好像只要测得够准，性质就能从物体里取出来。

物理学成熟的地方不在于测得更准，而在于它一直追问：**当观察方式、参考系、尺度或者外部条件变了，哪些关系还保持，哪些量会跟着变？**

速度就是最简单的例子。同一辆车，路边的人看它是一个速度，并排行驶的车上的人看它是另一个速度。速度不是贴在车身上的标签，是车和参考系之间的关系。真正有物理意义的说法，不是孤零零一句“它的速度是 60 千米每小时”，而是说明相对于哪个参考系、在什么时间段、用什么方式测出来的。

这件事对数学学习有直接的用处。

一个孩子说“平移不改变图形的形状和大小”，如果他只是记住这句结论，“不变”就还是口号。只有当他能说清楚哪些点的位置变了、对应点的位移之间是什么关系、距离和角度为什么保持、哪些性质**不在**“不变”的范围里，他才真正知道这句结论是由什么撑着的。

同样，孩子背下“等式两边同时加上同一个数，等式仍然成立”，不等于他理解了等式的性质。撑着这条性质的不是“老师允许这样移项”，而是**等号两边表示的是同一个量，对同一个量做同一种可逆的操作，相等关系被保持**。

把“同时加”换成“同时乘”，撑着它的结构就多出一道边界：乘以零虽然保持相等，却不可逆，没法再除回去。换成“两边同时除以一个含未知数的式子”，还得先处理这个式子可能等于零。

**但物理学不允许我停在“抓住不变量”这句话上**。物理学自己的发展恰恰说明：一条规律不会因为在某些变换里保持，就自动获得无限的适用权。牛顿力学在日常速度和宏观尺度上极其有效，却不能无条件替代相对论；胡克定律在弹性限度内成立，超过材料的线性区域就变形甚至失效。

所以成熟的物理知识至少包含三件事：**什么变了；什么在变化中保持；这种保持在什么范围之外不再成立**。

数学教学长期只拿走了第二件。我们喜欢让孩子概括规律、找不变量、总结模型，却很少要求他说出：这条规律靠哪些条件托着，它从哪儿开始不再承重。

一个孩子会说“正比例函数的图像是一条过原点的直线”，可是当现实数据只在某一段近似线性时，他会把直线无限延长；会说“三角形内角和是 $180^\circ$ ”，却不知道这句话依赖的是欧几里得平面；会用“小数点对齐”做加法，却说不清对齐的不是符号的位置，而是**相同的计数单位**。

他记住了物理学最耀眼的部分——不变量，却丢掉了物理学最谨慎的部分——适用域。

## 心理学：表现越顺，有时越看不出学习发生了没有

**先说清楚这里用的是心理学的哪一块。**

是学习与记忆研究里一个已经很扎实的区分：**表现和学习不是一回事**。

- **表现**：此刻能观察到的行为——他现在能做对多少、多快。
- **学习**：相对持久的变化——过一段时间还在、换个场合还管用。

这两个经常一起走，但也经常反着走。有一类实验的设计特别值得知道，因为它简单到可以复述：

把学生分成两组学同一份材料。甲组反复阅读，乙组读一遍之后反复做**回忆测试**（合上材料，尽量把内容说出来，再核对）。

- **刚学完就测**：甲组更好——反复读的人当下更流畅；
- **隔两天或一周再测**：乙组明显更好。

这个现象叫**提取练习效应**。它带出一个更大的说法，心理学家管它叫**合意困难**：**在学习过程中制造一些恰当的困难，当下的表现会变差，长期的保持反而更好**。

还有一个更反直觉的结果：即使第一次回忆注定失败——你根本想不起来——只要随后给出正确答案，这次失败的尝试仍然可能让后面记得更牢。

对课堂的用处是这一句：“**讲完就练、练得很顺这件事，正是学习与表现最容易分家的地方**。材料刚读过、答案就在同一页、例题结构一样——这时候的流畅，很大一部分不是他的。

课堂上最容易观察的是表现。孩子答得更快，错得更少，练到第五遍几乎不用停顿。老师看见这种改善，自然把它当作学习。

心理学期区分这两件事。**表现**是当下可观察的行为，**学习**是相对持久、能支撑延迟保持和迁移的变化。两者经常一起出现，但不相等。

有一类经典实验是这样的：一组学生反复阅读材料，另一组反复做提取练习。短时间之后测验，反复阅读的一组表现可能更好；但两天或一周之后再测，做过提取练习的一组保持得更好。**训练当下最流畅的那种条件，不一定带来最持久的学习**。

还有研究把这一点推得更远：即使最初的提取尝试注定失败，只要随后拿到正确答案，这个“先尝试后学习”的过程仍可能促进后续记忆。

这些常被概括成“适当的困难有利于学习”。但更要紧的不是困难，而是**支架和承重的分离**。

材料刚读过、答案就在同一页、例题结构高度相似的时候，外部环境承担了大量提取工作。孩子表现很好，却未必需要自己重建关系。等到书合上、时间过去、说法变了，外部支架不再替他承重，原本藏起来的差异才会露出来。

这解释了课堂上一个特别常见的错觉：老师刚讲完一道题，马上让学生做一道同类练习，成功率很高。**不是因为学习已经稳定，而是题目的表面结构、老师刚才用的词、黑板**

上还没擦掉的图、短时记忆里的操作顺序，一起搭成了一座临时脚手架。孩子站在脚手架上完成了任务，我们却把整座建筑记在了孩子名下。

心理学给的这条线索很硬：**要判断学习发生了没有，必须改变支架，而不只是换数字。**

## 工程学：机器转了一次，不叫设计成功

先说清楚工程学在这里的两个基本动作。

**第一个：把“做对了”和“能用”分开。**

工程上有两个词长得很像，意思完全不同：

- **验证**：这个产品是否符合我们写下的技术要求？（图纸上要求承重两吨，它承得住吗）
- **确认**：这个系统在真实的使用环境里，满不满足使用者的需要？（承得住两吨，但用户其实每天要过三吨的车）

一个东西可以完美通过验证，同时彻底通不过确认。航天工程的手册里把这两件事分得很清楚，还要加一层**资格测试**：在超出正常水平的环境里试——更高的温度、更强的振动、更坏的情况——看它还剩多少余地。

**第二个：问功能是由哪个结构承担的。**

一座桥把车的重量传到桥面、梁、支座、桥墩、地基。你不能指着其中任何一个说“是它扛住了这座桥”。工程师画的是**载荷路径**：力从哪里进来，经过谁，最后到哪里去。

这里有一个1934年的发明特别值得讲给非工程专业的人听，因为它几乎就是课堂的写照。当时的电子放大器有个大麻烦：元件的性能会漂移，输出跟着不稳。一位工程师想出的办法叫**负反馈**——把输出的一部分反过来送回输入端做抵消。

结果是：**放大倍数变小了，但输出稳定得多，失真小得多，而且元件本身的波动几乎看不出来了。**

请注意最后半句。**输出稳定，不代表元件变好了；稳定是那个回路制造出来的。**

把这句话搬进教室：一个孩子连续答对十道题，可能是他内部的关系稳了，也可能是题型重复、关键词固定、老师巡视答疑、同桌提醒、软件逐步提示——**这一整套环境，正在像负反馈一样，吸收掉他理解上的波动。**

一台机器在实验台上正常启动，远远不等于它能投入使用。

工程学必须区分好几个层次：某个部件不符合技术要求，整个系统满不满足用户需要；正常工况下能不能跑，高温、振动、误差和局部故障下还能不能保持必要功能；单个模块测试通过，集成之后接口可不可靠。

航天工程里把“验证”和“确认”分得很清楚：验证问的是产品是否符合规定的要求，确认问的是系统在预期的使用环境里满不满足使用者的需要。一个系统不会因为“照图纸做出来了”就自动成为“能完成任务的系统”。

工程学在这里比学校评价多问了一层：**这个功能，究竟由哪个结构承担？**

一座桥把车辆的载荷传到桥面、梁、支座、桥墩和地基，任何一个局部构件都不能单独声称“是我扛住了整座桥”。电路的稳定可能来自负反馈；一个软件服务的持续运行可能来自冗余和故障转移。

有一个工程史上的事实特别值得带进课堂：**负反馈放大器可以主动牺牲局部增益，换来整体稳定、失真更小、元件波动的影响更弱。**输出看起来更稳定，不等于每个元件本身更稳定——稳定性是回路共同制造出来的。

把这句话搬回教室，会得到一个让人警醒的对照：

一个孩子连续答对十道题，不一定说明他内部的关系已经稳定。稳定可能来自题型高度重复、关键词固定、老师巡回答疑、同桌提醒、作业软件逐步提示，以及答案格式的约束。正如负反馈能吸收元件的波动，教学环境也能吸收孩子理解上的波动，让输出长期保持整齐。

但工程学同时限制了那种粗暴的做法——“把帮助全撤了才测得出真水平”。在工程里，随便拔掉安全装置不是诊断，是事故。测试必须规定环境、极限和停止条件。教育里也一样：不能为了看“真实能力”，突然撤掉必要的语言辅助、无障碍工具、计算器支持或者合理的教师帮助。对某些孩子来说，这些不是替他思考的脚手架，而是他进入问题的基本接口。

工程学真正教给我们的也不是“把题出难一点”。难度只是载荷的大小。工程更关心载荷的路径、接口、裕度和失效方式。同样一道难题，可以只是增加计算量，也可以真正改变承重关系；同样一次失败，可以说明孩子没基础，也可以说明某个表象支架被撤走之后，载荷没有新的去处。

## 它们打架的地方

三门学问的词，一张表收好。

| 这个词   | 一句话解释              | 在这一章里对应什么   |
|-------|--------------------|-------------|
| 参照系   | 你站在哪儿量它            | 这道题默认以什么为准  |
| 不变量   | 换一种变换之后仍然不变的量      | 条件变了仍然成立的关系 |
| 适用范围  | 这条规律从哪儿开始不再成立      | 失效点         |
| 表现／学习 | 此刻做得出来／过后还在        | 当堂对了／真的会了   |
| 提取练习  | 合上书自己回忆，比反复读更持久    | 撤掉支架再做一次    |
| 验证／确认 | 合不规格／能不能用          | 步骤对不对／会不会用  |
| 载荷路径  | 力从哪儿进、经过谁、到哪儿去     | 这次推理由谁承担    |
| 负反馈   | 用输出反过来抵消波动，牺牲增益换稳定 | 环境吸收掉了孩子的波动 |

三家看起来都在说“要通过变化、延迟和复杂任务来检验学习”。但它们对“凭什么算证明了有能力”给的答案，互相不服：

- **物理学说**：能力显露在变换中的不变量里。
- **心理学说**：能力显露在支架淡出之后的持久提取里。
- **工程学说**：能力显露在真实约束下的系统承载里。

三条不能互相替代。一个人可以认出不变量，却无法独立提取；可以长期记得，却进不了真实任务；可以借助工具让系统跑起来，却说不清为什么成立。

它们争的其实是同一件事：一次成功，究竟该记在规律名下、学习者名下，还是整个系统名下？

而三家共享着一个都没说出口的想法：只要输出在某种变化中保持稳定，就可以把撑住这种稳定的能力，归到某一个焦点对象上。物理学容易归给定律，心理学容易归给记忆结构，工程学容易归给系统设计——**学校最干脆，直接归给学生。**

三家自己的材料却都在拆这个想法。不变量必须连着变换结构才有意义；心理表现由线索、时间和提取条件共同产生，而且提取本身还会改变记忆；工程输出由反馈、冗余、接口和环境共同承载，局部部件的能力没法从系统输出里直接分离出来。

所以一次稳定的表现，既不只属于公式，也不只属于孩子，更不只属于题目。

它属于一套关系。而我们一直缺的，正是一个能说清“这次结果由什么承重”的东西。

#### 四、承重结构，和那张格子

我把这套东西叫作**承重结构**。

它不是孩子脑子里的一张知识网，也不是题目本身的逻辑结构。它是一个**数学结论成立的时候，孩子、问题和外部支架之间实际形成的那张承重网**。

它至少有五根柱子。

**第一根：对象。**现在把什么当成一个东西在研究？分数问题里，是把整条绳子当作“1”，还是把若干小段当作单位？函数问题里，研究的是一个数值、一条对应规则，还是一组变化关系？**对象选错，后面每一步都可能算得很对，却承载着错误的问题。**

**第二根：单位。**哪些量能进同一个关系，取决于它们以什么为一份。小数加法对齐的是相同的计数单位；平均速度里，路程和时间怎么组织决定了“平均”是什么意思；概率里，结果能不能等分成概率单位，取决于等可能，而不只是结果的个数。

**第三根：前提。**这条结论靠哪些定义、条件、定理，还有哪些没说出口的假设撑着。“两边同除以 $x$ ”需要 $x$ 不等于0；把样本比例当成总体概率，需要抽样方式和样本量的条件。

**第四根：表示。**图、表、语言、符号、工具，各自承担了哪些识别和推导的工作。有时候线段图让“总量一部分”关系显出来；有时候坐标系把几何运动变成代数变化；**有时候一张公式表直接替孩子完成了概念选择。表示从来不是中性的包装。**

**第五根：支架。**老师的提问、例题的顺序、同桌的提醒、计算器、AI的提示、教材的章节安排、反馈机制——它们正在替孩子完成什么。

支架不等于作弊。任何复杂学习都需要支架。**真正的问题是：孩子知不知道这些支架承担了什么，当支架变了，载荷有没有新的去处。**

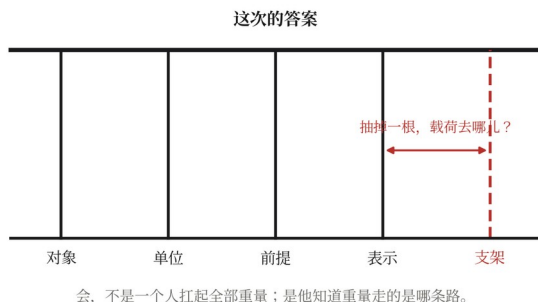


图3 一次正确的载荷路径：五根柱子，抽掉一根，重量往哪儿去？

有了这张承重网，还不等于学会了——一段程序在固定支架下也能形成稳定的承载。

真正的学习发生在**载荷能够转移**的时候。所谓载荷转移，不是把同一种方法搬到另一道相似题上，而是当某根柱子失效时，孩子能做到三件事：

1. 指出原来是哪一项在承重；
2. 说明它为什么在新条件下不够用了；
3. 把推理的任务转交给新的对象、单位、表示、前提或关系。

回到那张格子。竖轴这一章换成一句更具体的话：**撑着这次正确的东西变了，还站得住吗？**

|       | 这次做对了                    | 这次没做对                        |
|-------|--------------------------|------------------------------|
| 载荷转得走 | 活的正确：撑着它的柱子换了，孩子能重新安排    | 正在发生：还没做对，但已经能指出原来靠什么、现在哪儿不够 |
| 载荷转不走 | 借来的正确：答案对，可撑着它的是题面、模板和提示 | 普通的不会                        |

"借来的正确"是这一章的靶子。它比第一章的"封死的正确"更难看出来，因为它的成因根本不在孩子身上——**是环境太好了**。

题面标签帮他识别，例题顺序帮他选择，工具帮他计算，没说出口的前提帮他保证合法性，老师和答案帮他验证。孩子确实完成了一次真实的推理，只是那次推理的大部分重量，不在他这一边。

这就是为什么有些孩子平时作业几乎全对，一到综合题就"突然什么都不会"；为什么老师讲过的题稍微换个说法，孩子就说"这个没学过"。**不是知识突然消失了，是载荷变了以后，原来藏着的承重结构露了出来。**

## 五、六个场面，六根不同的柱子

(以下场景为构想，题目均为常见题型。)

### 场面一：题面的词在承重

一年级孩子看见"还剩多少"就做减法。他能独立做完十道同类题，却未必知道减法为什么成立。**"还剩"这两个字已经替他完成了关系判断。**

怎么查：把题目改成"原来有一些，用掉8个，现在还有5个，原来有多少"。用掉"还在，但这次要加。

### 场面二：措辞在承重单位的选择

四年级孩子见到"每人分得同样多"就做除法。他会算，但未必知道除法里有两种完全不同的单位组织：已知总量和份数求每份，和已知总量和每份求份数。

怎么查：把"12个苹果平均分给3个人"和"12个苹果，每人分3个"放在一起问，让他说清这两次的"1"分别是什么。

### 场面三：图上画出来的线在承重

求三角形面积，图里没画高。孩子说“我不会，因为缺少高”——说明那条画出来的线段承担了全部载荷。

如果他能把载荷转交给“顶点到底边的垂直距离”，自己作出高，或者通过等积变形拿到同一个距离，结构就转移了。

### 场面四：符号消去在承重

孩子解  $(x-1)(x+2)=0$ ，用零积定理得到两个解，很好。

现在把题目改成  $(x-1)(x+2)\div(x-1)=0$ 。如果他直接把  $x-1$  约掉，只写  $x=-2$ ，说明符号消去这个动作承担了载荷，而定义域的条件没有进他的结构。真正的载荷转移，是在约分之前就把“ $x$  不等于 1”当作一根柱子留下来。

这是初中最值得反复用的一道诊断题。它不难，但它精确地测出一件事：孩子知不知道自己刚才那一步是靠什么合法的。

### 场面五：章节标题在承重

初中生见到一次函数就找  $k$  和  $b$ ，见到反比例函数就写  $xy=k$ 。他会套式，却可能不知道这个模型成立的变量范围和现实条件。是教材的章节标题替他完成了模型分类。

怎么查：给一组数据，不告诉他这一章在讲什么，问他“这两个量之间可能是什么关系，你凭什么这么说”。

### 场面六：老师的即时反馈在承重

这一个最隐蔽。孩子每写一步，老师就“嗯”一声或者皱一下眉；孩子据此调整，最后写出完全正确的解答。

整个过程里，没有一条错误关系真正被他自己重建过。他学会的是读老师的脸。

怎么查：把反馈延后。让他先把整道题走完，再一起看。

### 六个场面合起来说明什么

这六次成功，撑着它们的分别是：题面的词、措辞、画出来的线、一个符号动作、章节标题、老师的表情。

六根柱子里没有一根是错的。题面本来就该有词，教材本来就该分章节，老师本来就该给反馈。问题从来不是“有支架”，而是我们把支架承担的那部分，一起记在了孩子名下。

## 六、明天就能用的一张卡

**第一，设计“支架扰动”，而不是单纯加难度。**

改题的时候先问自己：我这次改的是计算量，还是承重结构？把数字从两位改成三位，是加载荷；把画出来的高撤掉，是换柱子。后者比前者便宜得多，也有用得多。

**第二，在答案之外，让孩子写三句话。**

- 原来那个方法靠什么成立？
- 这次的变化，让哪一步不再够用？
- 现在由什么接手？

这三句是整本书里性价比最高的三句。它们不增加做题量，只增加一次说明。

### 第三，区分“用了提示”和“依赖提示”。

用了提示不是问题。要看的是：提示撤掉以后，那部分工作有没有去处。同一个孩子在同一道题上，用提示做对和不用提示做对，是两种不同的结构状态——**但它们在成绩册上是同一个勾。**

### 第四，延迟任务要用结构变式，不要用同构复现。

两周后再考一道数字不同、结构一样的题，测不出什么。要换的是那根柱子。

### 第五，评价结构位置，不评价人品。

“借来的正确”说的是这一道题上的位置，不是“这个孩子不独立”“这个孩子懒”。位置随时会变，标签不会。

### 第六，老师也给自己画一张载荷图。

上完一节课，问自己一句：这节课里，孩子的哪一部分推理，其实是我替他做的？——**这一问会比任何听课评价都更准确地告诉你这节课发生了什么。**

## 七、这套眼光的分寸

### 第一，撤支架不是诊断手段，是设计动作。

工程里随便拔掉安全装置不是诊断，是事故。教室里也一样。撤支架要有目的、有范围、有停止条件：撤哪一根、撤多久、孩子卡住到什么程度就接住他。**为了“看出真水平”而突然把帮助全撤掉，测出来的往往只是慌乱。**

### 第二，有些工具不是脚手架，是接口。

语言辅助、无障碍工具、对某些孩子而言的计算器，不是替他思考的东西，是他进入问题的通道。撤掉它们不叫暴露真实能力，叫把人挡在门外。**这条没有例外。**

### 第三，难度和载荷是两件事。

一道题很难，可能只是算起来麻烦。承重结构一点没变的难题，做一百道也测不出这一章说的东西；而一道很简单的题，把一根柱子抽掉，立刻就能看见。

### 第四，支架多不是罪。

这一章不是在说“帮助越少越好”。一个系统没必要拆掉所有反馈来证明零件足够强壮。真正要避免的只有一件事：**支架承担了大量工作，而我们和孩子都不知道它承担了什么。**

### 第五，五根柱子不必每次全查。

不同内容有不同的薄弱柱。低年级多半是“题面的词”和“单位”，初中多半是“前提”和“表示”。一节课盯住一根，比五根都问一遍有用。

## 这一章的一句话

**“会”不是一个人独自扛起了全部重量。“会”是他知道重量走的是哪条路——所以当一根柱子被抽掉，他有地方把它放下。**

下一章我们把时间轴拉长，问第三个问题：学了新的以后，旧的那些，会不会被改写？

## 第三章·旧知识会不会被改写

### 第一编 什么才算真的学会·之三

#### 一、一个六年级孩子说的话

下面这个场面是我构想的，但那句话我相信你听过很多次。

六年级，分数乘法。老师问："乘法是什么意思？"

一个孩子举手，答得又快又准："乘法就是几个相同的数加起来。"

这句话在二年级是对的，而且是漂亮的对——他当年就是靠着它，把乘法从加法里长出来的。

问题是现在是六年级。

黑板上写着  $2/3 \times 1/4$ 。

按他那句话，这是" $1/4$ 个 $2/3$ 相加"。四分之一东西怎么相加？他自己也说不通。于是他做了一件几乎所有孩子都会做的事：**把那句话放在一边，另外记一条规则——分数乘分数，分子乘分子，分母乘分母。**

现在他脑子里有两样东西。一样是二年级那句"几个相同的数加起来"，还立在原地，一个字没改；另一样是六年级新记的一条操作规则，和前面那句没有任何关系。

他两样都会用。整数乘法用第一样，分数乘法用第二样。考试全对。

**但是他的"乘法"已经裂成两半了。**而且他不知道。

再往后走，负数乘法会裂出第三块，字母乘法会裂出第四块，向量点乘会裂出第五块。到高中，他手里握着一把互不相干的"乘法"，谁也不知道它们凭什么共用一个符号。

现在换一个孩子。同样的问题，他停了一下，说："以前我说乘法是几个相同的数加起来，那是因为那时候份数总是整数。现在份数可以是 $1/4$ ，'加几次'就说不通了。**得换个说法：乘法是求一个数的几分之几是多少——原来那种整数的情况，正好也包含在里面。**"

这句话里发生了一件很要紧的事：**他不是学会了一条新规则，他是回过头去，把二年级那句话改写了。**

改写之后，旧的没有被扔掉——整数乘法仍然可以看成连加，只不过它现在是一个特例，不再是定义。

这一章要讲的，就是这件事。

#### 二、这一章要还的债

**新知识学进去了，不叫学会。真正的学会，是新东西反过来把旧东西改写了一遍。**

这不是"温故知新"那种老话。温故知新说的是旧的帮助你学新的；这里说的是相反的方向——**新的回头去改旧的**，改完之后，那个旧知识的意思、边界和用法都不一样了。

#### 三、三门学问，三个"身份藏在哪里"

这一章我请来的三门学问，争的是同一件事：**凭什么说一样东西"还是原来那样东西"？**

**细胞学：一个神经元，必须不断做使自己继续是神经元的事**

**先说清楚细胞学在这里问的是什么。**

一个成年人身上大约有几十万亿个细胞，其中绝大多数带着**完全相同的一套DNA**。

那么问题来了：同一套图纸，怎么会造出神经元、肌肉细胞、皮肤细胞这些形态、寿命、功能差得极远的东西？

答案是：**DNA不是图纸，它更像一整本菜谱**。每个细胞都拿着同一本菜谱，但今天做哪几道菜、火开多大，由另一套东西决定——哪些基因被读取、什么时候读、读多少。管这件事的叫**转录因子**，它们像一组开关，成组地打开或关上某些页。

关键的一点在这里：**成熟细胞的身份不是一次定下来就锁死的**。很多成熟细胞必须**持续地**表达那几个关键的调控因子，才能维持自己是什么。把这套维持网络干扰掉，有些细胞会转成别的类型。

所以一个神经元不是因为“它曾经变成过神经元”，就永远是神经元。**它必须不断地做那些使自己继续成为神经元的事**。

有一个实验把这件事推到了极处。2006年，山中伸弥的团队发现：只要把**少数几个**转录因子送进一个已经成熟的皮肤细胞，就能把它“推回”到接近胚胎干细胞的状态——重新拥有变成多种细胞的能力。这就是常说的**重编程**，后来拿了诺贝尔奖。

它推翻的是一个很牢固的常识：**分化是不可逆的**。不是的。成熟的身份可以很稳定，但不是绝对封死；当调控网络被系统性改写，细胞会重新打开曾经关上的可能。

还有一件事对我们特别有用：细胞里的蛋白质、RNA一直在被合成和降解，**具体的分子来了又走，细胞却保持着同一个类型**。它保持的不是同一批材料，是**组织材料更新的方式**。

人身上大多数细胞携带着几乎相同的基因组。可是神经元、肌肉细胞、胰岛细胞和皮肤细胞，形态、功能、寿命差得极远。

如果细胞的身份能从DNA序列里直接读出来，同一套DNA就不该长出这么多种细胞。

细胞学早就把“基因是蓝图，细胞照图施工”这个想象变得不够用了。哪些基因被表达、什么时候表达、表达多少，由基因调控网络、染色质状态、转录因子的组合、细胞之间的信号、代谢状态和组织环境共同决定。

更要紧的是：**成熟细胞的身份不是一次分化完成后就自动冻住的**。许多成熟细胞需要持续表达关键的调控因子，才能维持自己的功能；干扰这套维持网络，有些细胞会转成别的谱系。

也就是说，一个神经元不是因为“它曾经变成过神经元”，所以以后永远是神经元。**它必须不断地做那些使自己继续成为神经元的事**。

这对“学会了就存在那儿”提出了第一重挑战。一个孩子曾经独立发现乘法分配律，不保证这条关系从此自动存在。它需要新的计算、面积模型、代数变形和函数表达里被反复激活、压制错误的竞争者、接上新的关系。

但细胞的维持又不是简单的重复运行。细胞里的蛋白质、RNA、膜成分持续地合成和降解，很多具体分子来了又走，细胞却保持着相对稳定的类型。**它保持的不是同一批材料，是组织材料更新的方式**。

这一条对数学学习特别重要。孩子今天解释比例，可以说“两个量同时扩大相同的倍数”；以后可以说比值不变；再以后可以用 $y=kx$ 、斜率、相似变换。旧说法被淘汰了，具体的例题早忘了，但某种关系组织仍在不断生成新的表达。

**如果我们坚持“记得原话才叫保存”，就会把真正的学习更新误判成遗忘。**

细胞学还给了一个更强的事实：已经分化的细胞能够被重编程——少数几个转录因子的组合，可以把成熟的体细胞诱导回多能状态。成熟的身份可以很稳定，却不是绝对封死的。

把这件事带进数学，会看见一个常被忽略的现象：**真正的学习有时候不是把旧知识加固，而是让旧知识失去它原来的身份。**

孩子最初把“数”理解成数物体数出来的多少；学了负数，数开始表示方向和位置；学了分数，数可以来自度量、比例和运算的结果；学了无理数，“写得完的小数”不再是数的边界。每一次扩展都不是把同一个“数”原样保存下来，而是**重写了数的规则**：哪些对象算数，什么操作被允许，什么关系成为核心，旧解释在哪个范围里还有效。

## 历史学：两座今天一模一样的城市，不是同一种城市

先说清楚这里用的是历史学的哪一块。

不是史实，是一种叫**路径依赖**的解释方式。它想回答：为什么有些明明不够好的安排，会一直用下去？

最有名的例子是键盘。你面前的键盘第一排是 QWERTY——这个排列并不是打字最快的排法，它当年是为机械打字机设计的（据说是为了让常用字母隔开，减少卡键）。机械打字机早就没了，可这个排列还在。

为什么？因为一旦大家都学会了它，改的成本就在不断上升：所有人都得重学，所有厂商都得改产线，所有教材都得重写。路径依赖研究把这叫**增益递增——一条路被走的人越多，改道就越贵；改道越贵，走的人就越多。**

于是历史学得出一个和“看当下就够了”直接对撞的判断：**最后的状态没法只用当前的变量解释，因为当前的变量本身就是历史沉淀下来的结果。**

有人用棋局做过一个很锋利的区分：

- 如果下一步只需要看**当前的棋盘**，那么历史信息已经全部写进了这一局面里，来路可以扔掉；
- 但如果**同样的棋盘**，因为形成过程不同而需要不同的下法，那么历史就还带着不能被压缩掉的信息。

这个区分直接可以用在孩子身上。两个孩子今天都会做三位数乘法，当前状态一样。但一个经历过对象的重新划分、单位的冲突、反例的修正，另一个只经历过范例模仿——**相同的当下，并不预示相同的未来。**

假设两座城市今天的道路宽度相同、人口相近、产业结构也差不多。它们是同一种城市吗？

历史学会先问：这些道路是怎么形成的？产业是战后重建、资源枯竭后转型，还是一次性由政策导入？看上去相同的当下配置，可能承载着完全不同的制度记忆和未来惯性。

路径依赖的研究把这件事说得很清楚：早期看似微小甚至偶然的**选择**，会逐步抬高改道的成本；一条路一旦被采用，就通过学习效应、协调效应和不断的投资自我加固。**最后的状态没法只用当前变量解释，因为当前变量本身就是历史沉淀下来的结果。**

有人用棋局做过一个很锋利的区分：如果下一步只需要看当前的棋盘，那么路径信息已经全部写进了状态；**如果同样的棋盘因为形成过程不同而需要不同的下法，历史就仍然带着不可压缩的信息。**

这对数学学习的意义，比“了解数学史有助于理解数学”深得多。

两个孩子今天都写出  $x=3$ ；两个孩子都会做三位数乘法；两个孩子都把圆面积写成  $\pi r^2$ 。当前状态一样。但如果一个孩子经历过对象的重新划分、单位的冲突、反例的修正和表示的转换，另一个只经历过范例模仿，那么**相同的当下并不预示相同的未来**。

形成的次序会改变知识的可用方式。先把分数理解成“把一个整体平均分成几份取其中几份”，再进入商、比率、测量和算子，跟一开始就把分数当成“两个整数中间画一横”，会形成不同的概念重心。先经历“为什么需要一个未知数”，再学方程，跟先背“设  $x$ 、列式、解答”的模板，会让同一个符号背负不同的历史。

**旧路径会以多种方式回到现在**：可能成为自动化的优势，也可能成为错误的吸引子；可能让人对某些表示特别敏感，也可能造成对另一些表示的盲区。

所以“遗忘”未必是把历史删掉了。孩子可能忘了某道例题，却留下了那次对象重划形成的倾向；忘了老师的原话，却留下一“遇到条件冲突先检查单位”的习惯。**历史不只以能讲出来的记忆存在，也以判断的偏向和默认的起点存在。**

**但历史学也有它的盲点**：如果把来路看成不可替代的身份来源，就容易把过程神圣化——好像孩子必须亲自重走一遍数学史才算真懂。这会忽略一件事：**知识是可以压缩、重编、重新进入的**。一个熟练的人不必每次做加法都重走一遍计数的历史。

## 化学：既相信身份，又允许构成身份的键不断交换

**先说清楚化学在这里怎么判断“还是不是同一个东西”。**

化学处理的全是变化：键断开、键形成、原子重新组合。它之所以能研究变化，是因为它必须先能辨认：什么在变，什么没变。

**第一层判据是组成**。水就是两个氢一个氧，不管这滴水是烧出来的、酸碱中和出来的，还是冰化的。**来源不改变它是什么。**

**第二层判据比组成更细：连接方式**。有一个词叫**同分异构体**——分子式一模一样，但原子的连法不同，于是变成完全不同的两种东西。

最好懂的例子是  $C_2H_6O$ 。同样的原子清单，可以连成两种东西：

- 连成一种，是**乙醇**——酒精，室温下是液体，能喝（虽然不建议）；
- 连成另一种，是**二甲醚**——室温下是气体，可以当燃料。

**同样的元素清单，不同的关系结构，就是两种物质**。所以化学的身份判据里，**关系结构必须进来**。

**第三层最有意思，也最新**。传统的强力塑料（比如热固性材料）之所以结实，是因为分子之间的共价键交联得死死的，不可逆。代价是：**它裂了就没法修，也没法回收重塑**。

近二十年发展起来的**动态共价化学**做了一件看起来矛盾的事：让一部分共价键在特定条件下（加热、光照、催化剂）**可以断开再接上**。结果是这类材料能自修复、能重新塑形、能回收，同时**整体的功能和连接仍然保持**。

一句话记住：**它之所以能保持更高层的身份，不是因为所有的键都拒绝改变，恰恰是因为一部分键能够在约束下交换**。

**稳定，不等于冻结**。

化学处理的就是变化。它之所以能研究变化，首先因为它必须辨认：什么在变、什么守恒、什么成了新的物质。

在经典的化学判断里，物质的身份主要看当前的组成、结构和可重复的性质。水分子是燃烧生成的、酸碱中和来的、冰化的还是蒸汽冷凝的，都不改变  $H_2O$  的化学计量关系。**来源不是完全无关，但它不能单独推翻当前结构给出的身份**。

这跟历史学直接对撞。历史学说来路不能被当下压缩；化学却必须在大量情形里把来路括起来，否则同一种物质会因为每条形成路径不同而被无限分裂。

化学还给了一个比“看组成”更细的判据：同分异构体的分子式一样，却因为原子的连接方式或空间排列不同而性质不同。**只看元素清单不够，关系结构必须进入身份。**反过来，两份样品即使不是同一批分子，只要组成、结构和反应关系等价，仍然算同一种物质。

这一条给数学学习的提示很有力：**身份既不等于材料相同，也不等于历史相同，而可能在于材料之间的组织关系。**

一个孩子用方格覆盖来理解面积，后来用底乘高，再后来用积分。具体对象、语言和算法全换了，但“把空间区域用可加的度量单位组织起来”这条关系可以在不同层次里连续下来。反过来，两个孩子都写底乘高，一个把“高”理解成斜边，另一个理解成两条平行线之间的垂直距离——**表面公式一样，关系身份根本不同。**

化学里还有一类材料把这件事推到了极处。传统的热固性材料靠共价交联获得稳定，代价是难以修复、重塑和回收。而动态共价网络通过可控的键交换，让网络在外界刺激下重新排列——**在保持整体连接和功能的同时，完成重塑、自修复和再加工。**

这里出现了一个看起来矛盾的事实：

**材料之所以能保持更高一层的身份，不是因为所有的键都拒绝改变，恰恰是因为一部分键能够在约束下交换。**

稳定不再等于冻结。

这跟数学学习太像了。一个知识如果只在所有符号、例题、步骤和说法都保持原样时才成立，它就像脆性的永久交联材料：局部极稳，整体无法重塑。题型一变，它不是缓慢调整，是直接断裂。

真正有生命力的关系，必须允许某些连接被拆开。“每份数×份数=总数”可以改写成“总量=单位量×单位个数”，再改写成矩形面积、匀速路程、单价总价、密度质量、函数斜率。**表面的连接一直在换，但每一次重组都必须受数量意义的约束**——不能为了追求灵活，就把所有乘法情境都硬归成同一个结构。

## 它们打架的地方

三门学问的词，一张表收好。

| 这个词   | 一句话解释              | 在这一章里对应什么        |
|-------|--------------------|------------------|
| 转录因子  | 决定哪几页菜谱被读取的开关      | 维持一条关系需要持续做的事    |
| 重编程   | 少数几个因子能把成熟细胞推回多能状态 | 旧知识可以失去原来的身份     |
| 路径依赖  | 走的人越多，改道越贵         | 先学的说法会一直影响后面     |
| 增益递增  | 越用越难改              | 二年级那句话为什么赶不走     |
| 同分异构体 | 原子一样，连法不同，就是两种东西   | 都写“底×高”，理解可以完全不同 |
| 动态共价键 | 可以断开再接上的强连接        | 拆得开的知识才活得下来      |

三家把“单独够用的证据”放在了三个不同的位置：

- **细胞学**放在持续运行的调控网络上；
- **历史学**放在不可压缩的形成次序上；
- **化学**放在当前可测的组成、结构与反应关系上。

谁也吞不掉谁。只按当前状态判断，就看不见相同表现背后不同的来路；只按历史判断，就会把知识的压缩和重新进入误判成身份丧失；只按当前结构判断，就会把会写同一个公式的人当成拥有同一条数学关系。

而三家共享着一个更隐蔽的想法：**数学知识的身份，总能从某个地方被读出来**。三家的争执，只是在争“正确的档案柜是哪一个”。

我的判断是：**根本没有那个档案柜**。

“比例”不是藏在某个定义、某段历史或某个应用里的实体。它在双数轴上被组织成两种单位的协同变化，在表格里被组织成对应值之间的不变量，在图像上被组织成过原点直线的斜率，在相似图形里被组织成对应长度的乘性关系。**每一次表达都改变了比例“是什么”，却又不能任意改变**。

如果比例只是一个永恒的内核、所有表达都只是外衣，那么新的表达不会反过来增加、限制或重写这个概念。可事实恰恰相反：函数图像会让孩子第一次看见比例里的整体趋势；单位分析会让他发现无量纲的比和有量纲的率不是一回事；反比例会逼他修正“一个量变大，另一个也按同样倍数变大”的直觉。

**关系不是先有了身份，再进入各种表达。关系是在表达的更换中取得身份的，也在更换中失去旧身份的一部分。**

## 四、代谢，和那张格子

我把这种组织叫作**代谢**——数学关系在材料不断更换的过程中，重新生成自身。

“代谢”在日常语言里常被理解成排出旧的、吸收新的。只取这个意思，它就退化成“更新知识”的花哨说法了。我用这个词，是要它带上五个不能拆开的动作。

### 第一，拆解。

孩子能把一个看起来完整的公式、步骤或图形拆成不同的角色：对象是什么，单位是什么，哪些是必要条件，哪些只是当前的表示，哪些操作保持关系，哪些步骤只为算得方便。

**不能拆解的知识没法代谢，它只能整块保存或者整块丢掉**。一个孩子把“速度和 $\times$ 时间=路程和”当成一个整体模板，只有当他能拆出“参照对象”“方向”“相对速度”“共同经历的时间”“起始距离”这些角色，这个公式才有重新组合的可能。

### 第二，保留约束。

代谢不是随便改写。旧形式拆开之后，必须认出哪些关系不能丢：等式变形可以换表达，但要保住解集；分数可以换分子分母，但要保住同一个数；图形可以剪拼，但要保住面积；统计图可以改纵轴刻度，但不能瞒住比较的基准。

**没有约束的“创新”，只是对象在漂移。**

### 第三，替换材料。

数字、情境、图形、语言、符号、算法顺序、工具或者协作方式里，至少有一样被真正换掉了，而且不是在原模板里填空。

把 $3+4$ 换成 $5+7$ ，不构成代谢。把“同单位计数”从小棒挪到钱币、长度，再挪到代数里的同类项，才开始触到材料替换。

### 第四，重新合成。

孩子在新材料里重建对象、单位和关系，而不是找旧题的表面对应。重新合成不保证第一次就对。关键的证据是：**他能指出新材料里哪一处没法被旧结构直接接管，并据此调整。**

**第五，回头改写旧的。**

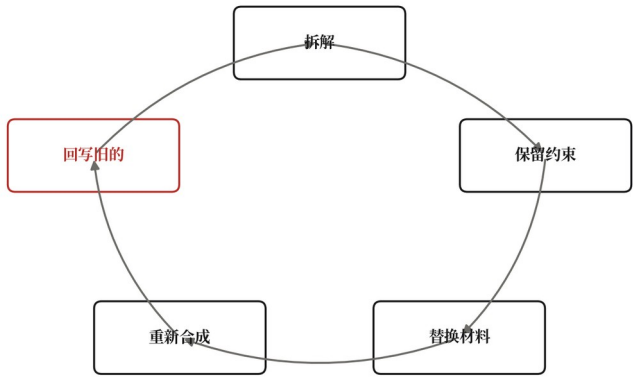
这是代谢和一般迁移最重要的差别。

新情境里的重组，不只解决了新题，还会反过来改变旧知识的意思。孩子从平行四边形面积回看长方形，会发现“底 $\times$ 高”里的高不是一条边，而是平行线之间的垂直距离；从负数乘法回看乘法，会发现“几个相同的数相加”再也覆盖不了全部乘法；从函数复合回看“代入”，会发现输入和输出的对象层级会变。

**如果新的应用成功了，却没有改写他对旧知识的理解，那更像是搬运；只有当旧知识在新合成之后获得了新的边界、角色或内部结构，这一轮代谢才算完成。**

所以：

**代谢，是一条数学关系在材料不断更换时，把旧关系拆成看得见的约束，扔掉偶然附着的东西，在新对象里重新合成，并用这次合成回头改写自己身份的能力。**



五个动作缺一不可。没有第五步，前四步只是变式练习。

图4 代谢的五个动作。没有第五步，前四步只是变式练习。

回到那张格子。竖轴这一次问的是：**旧的那些，被改写了吗？**

|         | 这次做对了                   | 这次没做对                   |
|---------|-------------------------|-------------------------|
| 旧的被改写   | 活的正确：新知识回头修改了旧知识的边界和意义  | 正在发生：还没做对，但旧知识的边界已经开始松动 |
| 旧的一个字没动 | 标本式的正确：新旧各占一格，各自都对，互不相干 | 普通的不会                   |

“标本式的正确”是这一章的靶子。标本很完整，形态漂亮，可以陈列——**它只是不再代谢了。**

本章开头那个孩子就在这一格。他的整数乘法是对的，分数乘法也是对的，两样都能得分。**它们只是互不相识。**

## 五、六个场面，六次身份改写

(以下场景为构想，题目均为常见内容。)

### 场面一：通分改写的不是分母，是"整体"

孩子会通分，会算  $1/3 + 1/4$ 。

把题目换成："一根绳子先用去全长的  $1/3$ ，剩下的部分又用去  $2/3$ ，一共用去全长的几分之几？"

如果他直接写  $1/3 + 2/3 = 1$ ，他的"分数"还是"上下两个整数"。要走通这道题，必须发生五步：把"分数=份数/总份数"拆开，认出**只有以同一个整体为基准的份额才能直接相加**这条约束，把第二个分数的整体从"全长"换成"剩下的部分"，重新合成  $2/3 \times 1/3 = 2/9$ ，最后回头改写"通分"的意思——**通分不只是找公分母，是把两个分数放回同一个整体**。

最后这一步才是代谢。前四步只是把题做对。

### 场面二：负数让乘法丢掉旧身份

就是开头那件事。

"几个相同的数相加"在  $(-3) \times (-4)$  面前彻底失效。孩子必须重新决定：乘法到底在做什么。是"求一个数的几分之几"，是"缩放"，还是"两次方向变换的复合"。

**这里不是给旧定义打补丁，是旧定义被降级成了一个特例。**这个降级动作，才是这一课真正要发生的事。

### 场面三：平行四边形改写长方形的过去

学平行四边形面积，通常这样讲：剪一个直角三角形补到另一边，变成长方形，所以面积是底乘高。

大部分课到这儿就停了。孩子学会了新公式，长方形那一课原样躺在原地。

再走一步：回头问一句——**"那长方形的'长'和'宽'，究竟是什么？"**

答案是：它们也是一对底和对应的高，只不过在长方形里，高恰好和一条边重合了，所以我们一直没注意到"高"是一个独立的东西。

这一问，把三年级的长方形面积改写了。改写之后，长方形、平行四边形、三角形、梯形就不再是四个要背的公式，而是同一条关系的四种情形。

### 场面四：方程不是解题路线，是一种关系身份

孩子最初把方程当成一条流水线：设  $x$ 、列式、移项、求解。

改写的时机在哪儿？在他第一次遇到  $3x + 5 = 3x + 5$  和  $3x + 5 = 3x + 6$  的时候（第一章我们从另一个角度看过这两道题）。这时候流水线还能开动，但产出的东西不再是" $x$  等于几"。

他必须回头改写"方程"这个词：**方程不是一条求  $x$  的路线，是一个关于"什么时候两个表达式相等"的陈述。**有唯一解只是其中一种情况，还有无穷多解和无解。

改写之后，"增根"就不再是计算事故——它是路线改变了解集。

## 场面五：函数图像不只是另一种画法

很多课把函数图像当成第三种表示方式：有解析式、有表格、还有图像。

如果它只是“另一种画法”，那它什么也没改写。

真正的改写是：**看过图像之后，“函数”这个词的意思变了。**在解析式里，函数看上去是一台代入求值的机器；在图像上，函数第一次成为一个**整体对象**——它有走向、有快慢、有对称、有极值。你可以谈论它的一段，谈论两个函数的交点，谈论把它整个平移。

这些话在解析式的框架里都很别扭，在图像的框架里才自然。**图像不是给函数换了件衣服，是让“函数是什么”这件事变了。**

## 场面六：证明改写的是“真”这个字

小学到初中，“这是真的”通常意味着：算出来是这样，量出来是这样，老师说是这样。

第一次做证明的时候，孩子经常抱怨：“这不是明摆着的吗，为什么还要证？”

这句抱怨恰好说明改写还没发生。

发生的时候是这样的：他忽然意识到，“**是真的**”和“**由什么使它成为真的**”是两回事。三角形内角和是  $180^\circ$ ，量一量确实是；但它之所以成立，靠的是平行线的性质。而平行线的性质靠的是平面这个前提。

这一改写完成之后，他回头看小学所有的结论，会发现它们各自都靠着某个没说出口的东西。**这是整个中学数学里分量最重的一次身份改写。**

## 六、明天就能用的一张卡

代谢不会自己发生。它需要一种课的设计。我把它压成四轮，一节课或者一个单元都能用。

**第一轮：原型轮。**让孩子在熟悉的材料里把关系稳定地建立起来。**这一轮不能省。**没有稳定的原型，后面换材料只会制造混乱。

**第二轮：单替换轮。**只换一样东西——一个数字变成分数，一个整体变成“剩下的部分”，一个图形从标准位置转过来。**只换一样。**换两样以上，孩子分不清是哪一处让旧方法失效的，代谢就变成了瞎猜。

**第三轮：重合成轮。**要求他用新对象重新建立关系，并且说出哪一条约束必须保住。这一步是防止“灵活”退化或乱改的闸。

**第四轮：回写轮。**回到原型，问一句：

“那我们上次说的那句话，现在还完全对吗？哪里要改？”

这一问是整张卡的核心。**没有第四轮，前三轮只是变式练习。**

一张四栏的记录（同样只记事实，不打分）：

| 栏     | 记什么            |
|-------|----------------|
| 拆出了什么 | 他把旧公式拆成了哪几个角色  |
| 保住了什么 | 他说出了哪条不能丢的约束   |
| 换了什么  | 这次真正被替换的材料是哪一样 |
| 改写了什么 | 他回头修改了旧知识的哪一句话 |

第四栏空着，就说明这一轮没有代谢，只有正确。

### 关于 AI，多说三句。

AI 最容易制造的，恰恰是“会自动更新的标准”——它随时能给出新材料下的正确解法，孩子一路照做，全对，旧知识一个字没动。所以用 AI 的时候加三道口子：

1. **问 AI 之前**，先让孩子写下“旧方法最可能在哪儿失效”；
2. **拿到多种方案之后**，让他删掉至少一种看着合理却破坏了约束的；
3. **关掉 AI 之后**，让他说出：哪一个词、哪个公式、哪条边界，因为这次对话被改写了。第三条是最关键的一条。它是判断这次人机对话到底发生了什么的唯一证据。

## 七、这套眼光的分寸

### 第一，不是每节课都要改写旧知识。

一个单元里真正值得改写的节点通常只有一两个：分数进来的时候，负数进来的时候，字母进来的时候，证明进来的时候，无理数进来的时候。**其余时间是巩固和熟练，不是代谢。把每节课都做成身份改写，孩子会失去落脚点。**

### 第二，记忆和自动化不是代谢的敌人。

能被代谢的关系，首先得是已经成形的关系。一个还没稳定建立起“份数与每份”的孩子，你去换他的整体，他不会重组，只会崩溃。**先有身份，才谈得上改写身份。**

### 第三，不要把孩子当成反复重编程的材料。

细胞可以被反复重编程，孩子不行。频繁的“打破你原来的想法”会造成一种特殊的伤害：孩子学会了不再相信自己上一次的判断，凡事等老师宣布。**代谢的目的是让旧知识获得新的边界，不是让孩子对旧知识失去信任。**

一个可靠的分寸是：**每一次改写，都要让孩子明确知道旧的那句话在什么范围内仍然是对的。**“几个相同的数相加”没有错，它在整数份数里一直有效——说清这一点，改写才是扩展，不是推翻。

### 第四，有些内容的价值就在于它不变。

定义、公理、符号约定需要共同遵守。代谢说的是关系在使用中重新生成，不是“什么都可以重新定义”。

### 第五，看不见第四栏，不代表没发生。

有些孩子的改写不通过语言表达出来，而是表现在他下一次的默认起点变了——上次遇到分数直接约分，这次先问“以谁为 1”。**这种沉默的改写同样算数，只是需要跨任务、跨时间才看得见。**

## 这一章的一句话

数学不是被保存下来的，是被不断重新合成的。判断一个孩子有没有真的学会一样新东西，最快的办法是回头看：他的旧东西，动了没有。

第一编到这里就完了。三章问了同一件事的三个侧面：**没走的路还在不在（第一章）、是谁在替他撑着（第二章）、旧的有没有被改写（第三章）。**三个问题合起来，才是“什么才算真的学会”的完整回答。

从下一章开始，我们换一个问题：**知道了要看什么，一堂课该怎么上，这些东西才长得出来？**

## 第二编 一堂课该留下什么

第四至六章

## 第四章·留下下一堂课绕不开的东西

### 第二编 一堂课该留下什么·之一

#### 一、一节很好的课，下课之后什么也没留下

下面这个场面是我构想的，但凡是听过公开课的人都会觉得眼熟。

四年级，一节讨论课。题目是："怎样公平比较两个小组的投篮成绩？"

甲组投了 20 个中 12 个，乙组投了 10 个中 7 个。

孩子们吵得很热闹。有人说比总进球数，甲组多；有人说比命中率，乙组高；有人说要看最好的一次；还有人说应该让他们投一样多的次数再比。

老师听得很欣赏，最后总结："大家说得都有道理。不过在数学上，比较这种情况我们通常看**命中率**，也就是进球数除以投篮数。"

孩子们点头。下课铃响。这节课评优，评委给了高分——参与度高，思维活跃，落点准确。

现在我们看第二天。

第二天的练习册上是一组新题：某商店两种商品的合格率、某班两次考试的及格率……全都是"求百分率"。孩子照着昨天学的做，全对。

**昨天那场争论，一点痕迹都没有留下。**

那句"应该比投得多的还是投得准的"，没有变成第二天任何一道题的条件；那个提出"让他们投一样多次数"的孩子，他的想法没有进入任何一个后续任务；那场热闹本身，只是老师抵达"命中率"这个结论的一段更生动的路。

我把这种课叫作**蒸发的正确**：课上确实发生了真实的思考，可是那节课结束的时候，思考也跟着结束了。它没有变成任何东西，所以下一节课可以完全绕开它。

#### 这件事有多普遍：一个来自录像研究的数字

我说"蒸发"，你可能觉得这是个修辞。它不是。

1999 年，国际数学与科学趋势研究做了一次大规模的课堂录像调查：七个国家的八年级数学课，一千多节，全部录像、逐字转录、按统一的规则编码。这是迄今为止关于"课堂上到底发生了什么"最扎实的一批材料。

研究者做了一件很要紧的区分。他们把课上的题目按**提出时的意图**分类，其中一类叫"建立联系"——就是那种本意不在于算出答案，而在于让学生看出两个东西之间关系的题。然后他们再看第二件事：**这道题最后在课堂讨论中，真的建立起那个联系了吗？**

结果是这样的：在澳大利亚，**平均只有 8%** 的这类题最后真的走到了建立联系；在美国，**不到 1%**。

请把这个数字读慢一点。

这些题**被提出来的时候是对的**。老师选了好题，教材编了好题，意图完全正确。它们**是在课堂运行的过程中被磨平的**——变成了套步骤、对答案、记结论。

这就是"蒸发"最准确的画面：**不是没有好东西发生过，是好事发生了，然后在这节课结束之前被抹掉了。**

顺带说一句同一份研究里的另一个数字：美国的课上，**69%** 的题目是"使用程序"型的。这不是哪个老师的选择，这是一整套系统的形状。

## 二、这一章要还的债

一堂课上得好不好，不看这节课结束时孩子会了什么；看下一堂课能不能绕开它。

如果下一节课少了这一节，照样能上——那么无论这节课多热闹、多精彩、多受表扬，它在学习上都没有留下东西。

## 三、三门学问，三个不同的“终点”

### 语言学：孩子不是先学语法，再开口说话

先说清楚这里用的是语言学的哪一派。

叫**使用本位**的语言习得研究。它和传统语法教学的分歧，正好可以用“孩子怎么学会说话”来看清楚。

一个两岁孩子最先说出来的不是语法规则，而是一些**整块的、连着场景的表达**：“妈妈抱”“还要”“不要这个”。他不知道“抱”是动词，也不知道这是个祈使句。这一整块，对他就是一个词。

接下来发生的事很有意思。随着相似的使用越来越多，他慢慢发现：

- 有些位置**可以替换**——“妈妈抱”里的“妈妈”能换成“爸爸”“奶奶”；
- 有些形式**承担相近的功能**；
- 有些顺序**会改变谁对谁做了什么**——“狗咬人”和“人咬狗”不一样。

于是一个可以生产新句子的框架，从一堆具体的使用里**长出来了**。语言学管这种带槽位的框架叫**构式**。

关键结论是这一句：**抽象的规则不是在使用之前等着被调用的模板，它是从大量具体使用中组织出来的**。而且频率会塑造规则本身——不只是让规则跑得更快。

但这一派同时给了一条很重的警告。频繁使用能长出规则，也能固化局部的习惯。孩子看见“比……多”就加，看见“平均”就除，看见两个分数就通分——**他不是没有从使用中学习，恰恰是学得太成功了**：某个课堂环境里的语言形式和操作反应，被牢牢绑在了一起。

所以“在情境中多用几次”并不自动产生数学关系。**使用可以长出规则，也可以长出套语**。

成年人学外语，最熟悉的顺序是先学语音、词汇、语法，再听说读写。这个顺序看起来很理性：没有规则，凭什么用得对？

孩子学母语不是这样的。

孩子最先拥有的不是一套语法，而是一些和情境、意图、交往目的缠在一起的整块表达：“妈妈抱”“还要”“不要这个”。随着相似的使用越来越多，他慢慢发现某些位置可以替换，某些形式承担相近的功能，某些顺序会改变谁对谁做了什么。

**抽象结构不是在使用之前等着被调用的模板，它是从大量具体使用中组织出来的**。高频使用不只是让一条已有的规则跑得更快，它会塑造规则本身。

这对数学教学是个直接的挑战。

教材通常先写定义、性质、公式，再让孩子通过例题学习使用。它默认一个概念可以先以清晰的形式存在，然后被应用。

使用本位的语言学会反问：**如果孩子从来没有在一次真实的交往里需要过这条关系，他学到的究竟是关系，还是一句可以复述的话？**

比如速度。老师先说“速度表示单位时间内通过的路程，速度=路程÷时间”，再给十道代公式的练习。孩子算得很熟，却未必经历过这样一种压力：两个人走的路程不同、时间也不同，**怎样用一个双方都认可的量，说服对方谁更快？**

只有当“快”这个日常判断不再够用、他必须让别人接受自己的比较时，“每一单位时间走多少”才不只是定义，而成为一种**表达上的需要**。

这里的“交往”不等于小组讨论，更不等于把生活故事套在算式外面。真正的交往有一个不能取消的条件：**必须有某个人，依靠你的数学表达去作判断、继续操作或者检验结果。**没有接收者、没有后续动作，所谓表达只是把老师心里的答案重新说一遍。

**但语言学也给了一条警告，而且很重。**频繁使用会组织出规则，也会固化局部习惯。孩子看见“比……多”就加，看见“平均”就除，看见两个分数就通分——他不是没有从使用中学习，恰恰是**学得大成功了**：某个课堂环境里的语言形式和操作反应被牢牢绑在一起。

所以“在情境中反复使用”并不保证数学关系真的发生。使用可以生成规则，也可以生成套语。

## 生态学：生物不只是适应环境，它还改造环境

**先说清楚生态学在这里挑战的是什么常识。**

常识是：环境是给定的，生物去适应它，适应得好的活下来。

**生态位建构**这个理论说：这只讲了一半。生物在活着的过程中会**改造环境**，而被改造过的环境，会成为后来者面对的新条件。

三个例子就够了：

- **海狸**筑坝，把一条溪流变成池塘，水位、植被、鱼类全变了；
- **蚯蚓**在土里钻来钻去、吞土排土，几十年下来土壤的结构、透气性、酸碱度都变了——达尔文晚年专门写过一本关于蚯蚓的书；
- **人**用火、农业、城市、制度和技术，把自己的生存条件改得面目全非。

更关键的是第二步：**这些改造会被继承下去**。下一代海狸不是在一条原始溪流边出生的，它一出生就在池塘边；下一代蚯蚓钻的是已经被上一代改良过的土。生态学管这个叫**生态继承**——它不通过基因传，通过环境传。

**这个理论也自带警告**：改造环境不天然是好事。生物造出来的环境可能提高适应度，也可能造成锁定和脆弱——把自己困在一条越走越窄的路上。

搬到课堂：一个班级完全可以共同建造一个非常稳定的**错误世界**——“答得快的人有话语权”“复杂问题就是猜老师想要什么”。所以“让学生共同建构”本身不能当答案。

学校最常见的学习模型是“适应”：课程规定学什么，教材安排顺序，老师选择任务和评价，孩子通过调整自己去适应。适应得快叫聪明，适应得稳叫基础扎实，能在陌生题型里继续适应叫迁移能力强。

生态学里有一个理论叫**生态位建构**，它指出生物并不只是在给定的环境里被动适应。通过代谢、迁徙、筑巢、挖洞、遮阴，生物会改变自己和别的生物所面对的选择压力。海狸筑坝改变水流和植被，蚯蚓改变土壤结构。**环境筛选生命，生命也持续重写环境。**

更要紧的是：被改造过的环境会被后来者继承。后面的行动者不是从零开始适应一个自然背景，而是在前一轮生命留下的痕迹、资源、风险和通道里继续行动。

**数学课堂经常切断这种继承。**

孩子今天发现一道题有两种合理解释，明天教材还用原来的表述；今天用一张自制的图表解决了问题，展示完就消失了；今天指出一个公式在某种边界下失效，后面的练习仍

把它当作无条件的规则；一届学生积累的典型错误、好问题和有效表示，不会成为下一届的起点。

**课堂允许孩子改变自己的答案，却很少允许孩子改变他将要学习的环境。**

这就造成一个奇怪的现象：学校不断要求孩子具备“改造世界”的能力，却把课程本身设计成一个不可被孩子的痕迹改变的世界。

数学课堂里的“环境”不是指教室的桌椅，而是指：哪些对象已经被命名；哪些表示可以直接调用；哪些工具允许使用；哪些错误被保留下来当警示；哪些问题被认为值得继续追问；哪些数据、模型和案例成了后续判断的公共材料。

回到开头那节投篮课。老师直接总结“应该比命中率”，环境没有被孩子改变——老师只是借他们的活动完成了一次更生动的讲解。

另一种做法是：把每一种比较规则写进班级的“比较工具库”，标明它在什么数据条件下会改变结论。**下一轮任务故意选一组数据，让总进球数和命中率给出相反的排名。再下一轮，让孩子自己设计一组能击穿某条旧规则的数据。**

这时候，昨天那场争论就不再停在“经历”层面，它成了今天任务的约束。**前一轮活动改变了后一轮的选择压力。**

**但生态学同样给了警告：**改造环境不天然是好事。一个班级可以共同建造一个非常稳定的错误世界——形成“答得快的人有话语权”的生态，把复杂问题简化成猜老师的意图，用同伴之间的暗号快速识别题型，把漂亮的板书当成理解的替代品。

**一群人可以共同建造一个稳定的错误世界。**所以“环境由学生共同建构”本身不能当答案。

## 建筑学：建成不等于完成

**先说清楚建筑学在这里的那个专门问题。**

一栋楼盖完、验收合格、交钥匙——工程流程到此结束，设计团队和施工团队就撤了。

然后人搬进去住。接下来的事情，图纸上都没有：夏天西晒到什么程度、走廊里说话听不听得清、实际能耗比设计值高出多少、大家会不会都挤在同一部电梯、那个漂亮的中庭冬天冷得没人待。

行业里有一个反复被证实的现象：**很多建筑的实际能耗，明显高于设计阶段的预测值。**原因往往不在楼本身，而在设备怎么被调、管理怎么做、人怎么用。

所以有一件事必须说清楚：**一栋楼的性能，不是这栋楼单方面拥有的属性；它是建筑、设备、管理和使用者共同作用的结果。**

针对这个断点，建筑学发展出**使用后评价**：在楼交付并被使用一段时间之后，回去实测、访谈、收数据，看它到底表现如何。还有一种做法叫**软着陆**——让设计和施工团队在交付之后继续跟进一段时间，把实际使用中的反馈带回调整，并且带进下一个项目。

一句话记住：**交付不是认识的终点，是这栋楼进入真实使用的起点。**最关键的性能知识，在验收之后才开始出现。

三门学问里，建筑看起来最相信结果。语言可以持续变化，生态可以长期演替，但一座建筑必须落成——图纸要变成墙体、楼板、门窗、管线，预算、材料、荷载、法规、气候和工期都必须在一个看得见的东西里结算。

这让建筑学对教育里大量“只有想法、没有承压物”的活动保持警惕：孩子说自己懂了，我们无法知道这种懂能不能被别人使用；孩子提出一种方法，却没有把条件、步骤和边界建成清晰的结构；孩子写出一份证明，读者必须不断猜他省略了什么。

建筑学的要求很简单：**把它建出来，让别人进去。**

一件数学造物可以是一张数据图、一个算法、一套分类标准、一份证明、一个可操作的模型、一份反例库、一道给低年级用的题。它的共同特征不是“动手做”，而是——**作者不在场解释的时候，仍然有人必须靠它完成某件事。**

**而建筑学最深的经验恰恰是：建成并不等于完成。**

传统的设计流程容易在交付的时候中断：设计团队完成图纸，施工团队完成建设，验收确认符合规定，然后各方离开。建筑真正被人使用之后，温度、声学、照明、能耗、动线、可达性和使用习惯才开始显露。这些问题在图纸和验收里看不出来，**因为建筑的性能不是这栋楼单方面拥有的属性，而是在建筑、设备、管理和使用者的共同作用中出现的。**

于是建筑行业发展出“使用后评价”，甚至有一种叫“软着陆”的做法：让设计和建设团队在交付之后继续参与，把实际使用中的反馈带回调整 and 后续项目。

这对数学学习提出了一条比“作品化学习”严格得多的要求：**孩子建成的东西，必须经过作者不在场、也控制不了的真实使用。**

作者站在旁边解释的时候，作品的缺口会被作者本人补上。只有作者离开，另一个人照着图操作、依着模型判断、按着证明复核，造物里哪些关系是真的建成了、哪些还只在作者脑子里，才会显露出来。

**但建筑学也会走到反面。**为了建成，必须冻结大量选择。一旦过度强调可交付的成果，学习会退化成作品主义：孩子花大量时间美化海报、做模型、排练汇报，却没有形成能生成新关系的组织。

## 它们打架的地方

三门学问的词，一张表收好。

| 这个词   | 一句话解释              | 在这一章里对应什么       |
|-------|--------------------|-----------------|
| 使用本位  | 规则从大量使用中长出来，不是先给再用 | 先让关系承担交付，再谈定义   |
| 构式    | 带槽位、能生产新句子的框架      | 能生成新解法的关系组织     |
| 套话    | 用熟了但只对熟悉措辞有效的固定反应  | 看见关键词就套方法       |
| 生态位建构 | 生物改造环境，环境再筛选生物     | 学生的活动改变下一轮的条件   |
| 生态继承  | 被改造的环境传给后来者        | 这一届留下的东西成了下一轮起点 |
| 使用后评价 | 交付之后回去实测它到底好不好用    | 把造物交给陌生使用者去用    |
| 软着陆   | 设计者在交付后继续跟进并回写     | 故障回写下一轮学习条件     |

把三家并排写出来很容易：语言学告诉我们多使用，生态学告诉我们创设环境，建筑学告诉我们做作品。

**如果结论只是“在真实情境中让学生合作制作作品”，那这一章就白写了。**

三家真正的冲突，在于它们把一条学习路径的**终点**放在了三个互不相容的地方：

- **语言学**把终点放在**生成规则**：一次表达只有被组织成能生产新表达的构式，才算学会。

- **生态学**把终点放在**条件的改变**：一次学习如果没有改变后续行动的工具、资源和限制，孩子只是适应了老师提供的场景。

- **建筑学**把终点放在**可使用的造物**：内部的理解和环境的改变都不能替代公共承压。

三条路平行摆着，老师立刻陷入困境：先重视哪一个？先抽象规则，可能过早关闭使用；先开放活动，可能固化局部生态；先做作品，可能让形式制作替代关系生成。

而三家共享着一个几乎从没被提出来的想法：**一条学习路径，存在一个可以宣布完成的终点。**

学校制度尤其需要这个想法。课时有结尾，单元要验收，作品要评分，学期要结算。所以即使是反对灌输的改革，也常常保留同样的结构：探究结束，老师总结；项目结束，学生展示；体验结束，概念形成；评价结束，进入下一单元。

**上一轮的产物可以被收藏、记录、表扬，却不是下一轮任务必须继承的条件。**

推翻这个想法的材料，不用从外面找，建筑学自己给了最硬的一击：**一座建筑在图纸、施工和验收层面全部完成之后，最关键的性能知识才刚刚开始出现。**交付不是认识的终点，是它进入真实使用的起点。

## 四、终点换手，和那张格子

我的判断是：**每一门学科追求的那个终点，都不能保留终点的身份。它一旦形成，就必须换手，成为下一条路径的起点。**

这就是我说的**终点换手**。整条路线是一个四步的循环：

**第一步：先用起来。**让一条关系先去承担一次真实的交付，而不是先被定义。

**第二步：把差异写进环境。**使用中暴露出来的分歧，被写进任务、工具、表示和班级的共同规则。

**第三步：建成别人能用的东西。**把关系外化成一个作者不在场也能运行的公共造物。

**第四步：让故障拿到改写权。**真实使用暴露的故障，回过头改写下一轮的学习条件。然后第四步接回第一步。

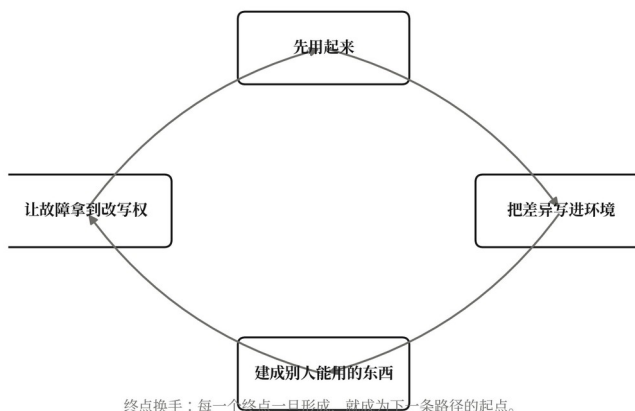


图5 关系栖造：终点换手，每个终点都成为下一轮的起点。

四步各自的关键，用四句话说：

**第一步不是创设情境，是制造必须交付的使用。**一次交付要成立，得同时具备三个条件：有一个接收者；接收者掌握的信息和你不同；表达失败会产生后果。三个条件缺一个，“情境”就只是包装。

**第二步不是教师总结，是把差异写进环境。**差异如果只停在讨论里，就随下课蒸发；写进工具库、写进下一轮任务的数据、写进班级的判断规则，它才留下来。

**第三步不是做作品，是建造“别人能够误用”的东西。**这句话的重点在“能够误用”——一件别人无法误用的东西，通常也是一件别人无法使用的东西。

**第四步不是反馈评价，是让故障取得改写权。**使用者在哪一步误用、停住、产生了不同的解释，这些必须具体地改变下一轮的某一个条件，而不是变成一句“下次注意”。

回到那张格子。这一章的竖轴是：**这节课有没有留下下一节课绕不开的东西？**

|           | 这节课很成功                     | 这节课不太顺                        |
|-----------|----------------------------|-------------------------------|
| 留下了绕不开的东西 | 活的正确：下一轮的起点被这一轮改写了         | 正在发生：课上得磕磕绊绊，但产生了一个下节课必须处理的分歧 |
| 什么也没留下    | 蒸发的正确：热闹、准确、被表扬，下一节课完全可以绕开 | 普通的不顺                         |

## 五、六个场面，六次“留下了东西”

（以下场景为构想，内容均为常见教学单元。）

### 场面一：分数单位是被交付失败逼出来的

不先讲“分数单位”。让每个小组给另一个小组写一份“我们用掉了多少”的报告，接收组必须据此算出剩下多少。

第一轮一定会出事：甲组写“用了  $\frac{1}{3}$ ”，乙组问“ $\frac{1}{3}$  是什么的  $\frac{1}{3}$ ”。

**这个问题不是老师提出来的，是交付失败逼出来的。**而“整体必须说清楚”这条规则，从此写进班级的报告格式——这就是第二步。

### 场面二：比例学习要让规则反过来去设计数据

先让孩子提出比较规则（比如“哪个更划算”）。然后反过来：“请你设计一组数据，让你自己那条规则给出明显不合理的答案。”

能设计出来，说明他知道自己那条规则靠什么成立。这组数据进入题库，下一轮别人做——造物建成了。

### 场面三：方程是别人可以复核的关系机器

不满足于孩子自己解对。要求他写一份“给同学的解题说明”，交给一个还没学过这一节的同学，只按说明操作。

哪一步被误解，哪一步说明不足，全部显露。**这比老师批改准确得多，因为老师会自动补全孩子省略的东西。**

#### 场面四：几何证明从“让别人无法误读”开始

证明的本质就是一件公共造物：作者不在场，读者必须能独立复核。

所以初学证明时，评价标准不该是“写得像不像书上的样子”，而是：**把它给另一个同学，他能不能不问你就完全读懂？**他卡在哪一句，哪一句就是缺了前提。

#### 场面五：函数的表示方式要成为可改造的公共环境

一个班可以维护一份“表示方式清单”：表格、图像、解析式、实际过程，各自在什么时候最好用、在什么时候会骗人（比如图像的纵轴刻度）。

这份清单不是老师给的，是逐节课长出来的，而且后面的任务会用到它。**这就是可以被继承的环境。**

#### 场面六：概率模型要接受真实世界的投诉

孩子做完一个概率模型，让它去预测一件真实会发生的事——明天有多少人带伞、这周谁值日、抽签的结果。

**然后必须回来对账。**对不上的地方就是“入住投诉”，投诉具体改写下一轮模型的哪一个假设。

### 六、明天就能用的一张卡

上课之前，把这六个问题问自己一遍。它们不需要额外的课时。

1. 谁会使用孩子的数学表达，并据此继续做什么？
2. 原来的表达，在哪一点刚好不够用了？
3. 孩子产生的哪一处差异，将被写进下一轮的任务、工具或规则？
4. 他们要建成什么东西，是作者不在场也能运行的？
5. 陌生的使用者会在哪一步误用、停住，或者产生不同的解释？
6. 这次故障，具体改变下一轮的哪一个条件？

第三问和第六问是这张卡的骨头。**只要这两问长期答不上来，无论课上得多好，都是蒸发的。**

再附一张教师插手点的小表——同一个动作，插得早或插得晚，效果完全相反：

| 阶段   | 这时候该做的           | 这时候最容易做错的      |
|------|------------------|----------------|
| 交付   | 制造信息差，让表达失败有后果   | 预先给出唯一的标准表达    |
| 凝结   | 提供对照，追问哪一处差异有区分力 | 把所有差异一次讲完      |
| 改写环境 | 保存痕迹，重排后续任务      | 让孩子的成果停在展示     |
| 造物   | 提供材料和真实的使用者      | 站在作者旁边替他补漏     |
| 使用故障 | 区分是使用者的问题还是造物的问题 | 只给一个分数或一句笼统的评价 |
| 回写   | 把故障写进下一轮的条件      | 把反馈封存在这一节课里    |

## 七、这套眼光的分寸

### 第一，它不取消必要的直接讲授。

有些东西就该讲：符号约定、术语、标准算法、安全的操作顺序。讲得清楚利落，是老师的专业。这一章说的是**哪些东西不该讲**——那些本来可以由使用逼出来的关系，讲掉了就没了。

### 第二，不要把每一道练习都变成公共工程。

一个单元里，能做成真实交付的通常只有一两处。其余时间是练习和熟练。**每节课都搞交付，孩子会疲惫，老师会崩溃，而且交付会变成新的表演。**

### 第三，被继承的环境不一定是对的。

班级建起来的工具库、规则、常用表示，可能整体是错的。所以这套做法必须配一个推翻通道：**任何一条写进环境的规则，都要留着“什么数据能击穿它”这一栏，而且要有人定期去击穿。**

### 第四，别人使用得好，不等于数学上正确。

一份被同学广泛使用的解题指南，可能好用但不严谨。使用是检验它能不能被别人独立操作，不是检验它对不对。这两件事要分开评。

### 第五，不能用“下一轮还要用”为理由无限延迟评价。

学期要结算，单元要验收。终点换手说的是**没有哪个节点拥有最终解释权**，不是取消所有节点。该给分的时候给分，该说清楚的时候说清楚。

## 这一章的一句话

一堂课真正的产物，不是这节课结束时孩子会了什么，而是下一堂课的起点被它改成了什么样子。

下一章讲第二条路：孩子已经动起来了，走着走着偏了——**那一刻，才是学习真正发生的地方。**

## 第五章·第一次走偏的时候

第二编 一堂课该留下什么·之二

### 一、他改对了，三天后又错了

下面这个场面是我构想的，但每个老师都经历过无数次。

五年级，解方程  $3(x-2)=2x+5$ 。

一个孩子写下第一步： $3x-2=2x+5$ 。

老师在旁边看着，马上说："括号前面是3，里面两项都要乘。"

孩子"哦"了一声，改成  $3x-6=2x+5$ ，接着往下算，答案正确。整个过程不到二十秒。

课后老师会觉得这次指导很成功：发现及时，纠正精准，孩子当场就对了。

三天后的作业里，出现  $2(x+4)=3x-1$ 。孩子写下： $2x+4=3x-1$ 。

同样的错。

我们通常把这解释成"没记住"或者"不认真"。但如果我们回到那二十秒去看，会看见另一件事：

**那次纠正，替他完成了三件他自己没做的事。**

第一，是老师发现了这里有问题——孩子自己没有察觉任何异常，他写完  $3x-2$  的时候是心安理得的。

第二，是老师判断了问题出在哪里——是分配律，不是移项，不是符号。

第三，是老师决定了怎么改。

孩子只做了第四件事：执行。

他学会的是"老师说要乘进去，就乘进去"。他没有学会的是：**走到那一步的时候，凭什么发现自己已经偏了。**

这就是我要讲的那一格：**导航式的正确**。有导航的时候，路一次都不会走错；导航一撤，人不知道自己在哪儿。

它的可怕之处和前面几种一样：短期指标全面变好——正确率上升，速度加快，孩子的体验也更好，因为他很少卡住。而且它会自我加固：因为提示有效，老师和 AI 都更愿意继续增加提示，孩子就更有机会练习自己定位。

### 二、这一章要还的债

**学习不是把路走对，是走偏的时候还知道自己在哪儿。**

一个孩子从头到尾没走错，可能是因为他知道路，也可能是因为一直有人在替他打方向。这两件事在结果上完全一样，在学习上完全相反。

### 三、三门学问，三种"怎么在不确定里往前走"

**诊断学：医生不从病名开始，从问题表征开始**

**先说清楚诊断学在这里研究什么。**

不是研究疾病，是研究医生的脑子是怎么运转的——这门研究叫临床推理。

它最重要的一个发现是：**新手和专家的差别，不在于谁记得的病多。**

新手的做法是“症状对应疾病”：看见一个关键词，从记忆里取一个病名出来。专家的做法完全不同——他先把病人那一大段叙述，**压缩成一份有区分力的描述**。这份描述有个名字，叫**问题表征**。

对照一下就明白了。同样是“胸口不舒服”：

- 「中年男性、突发、活动时加重、压榨感、伴出汗」
- 「年轻女性、餐后灼痛、平卧加重」

这两句不只是信息量不同，**它们是两张完全不同的地图**——各自指向的可能病因几乎不重叠。专家做的第一件事，就是把杂乱的叙述变成这样一句话。

接着是第二件事：**鉴别诊断**。专家手上从来不止一个候选，而是同时挂着两三个可能，然后专门去找那种**能把它分开的证据**——做哪一项检查，两个可能会给出不同结果？

医生还有一套东西叫**疾病脚本**：把某个病的典型人群、诱因、发展过程、典型表现打包成一个可以快速调用的模式。

**但诊断学最诚实的地方，是它承认自己有一个致命毛病，还给它起了名字：过早闭合。**

意思是：某个诊断解释了大部分现象之后，人会自然停止搜索——哪怕还有别的解释同样说得通。相关的还有**锚定**：最先想到的那个判断，会把后面所有信息都往它那边拽。

**最危险的不是没有候选，是某一个候选过早地获得了免于继续受审的地位。**

病人说：“胸口不舒服。”

这句话不是诊断，甚至还不足以开始推理。哪个位置？是压榨感、刺痛还是灼热？持续多久？和运动、进食、体位有什么关系？有没有伴随出汗、气促？

初学者容易把临床推理理解成“症状对应疾病”：看到一个关键词，从记忆里提取一个病名。而经验丰富的医生往往先把大量信息**压缩成一份问题表征**——把具体的叙述转化成有区分力的特征。

“中年男性、突发、活动相关、压榨性胸痛、伴出汗”和“年轻女性、餐后灼痛、平卧加重”，不只是信息多少不同，而是**把同一种日常表达组织成了两张不同的地图**。

这对数学教学的第一个冲击是：

**题目不是问题。题面上的句子、数字和图形，只是一组还没有被诊断的征象。**

传统教学常说“先读懂题”。所谓读懂，往往等于能复述已知条件和所求。诊断学会追问：这些条件为什么值得保留？**哪一个差异能把两个候选关系分开？**哪些信息只是噪音？如果删掉某个条件，这道题的身份会不会改变？

拿一道行程题：甲车每小时 60 千米，乙车每小时 40 千米，两地相距 300 千米，同时出发相向而行，几小时相遇。

只做信息摘录的孩子写下：甲 60，乙 40，总路程 300，求时间。

真正的问题表征至少要完成三次压缩：两个人消耗的是**同一段间距**；每过一小时，间距减少的量由两个方向共同决定；“相遇”不是两段路程分别达到某个数，而是**剩余间距变成零**。

这三条一成立，题目才从一串数字变成一个关系判断。

**而且诊断学不允许只有一个候选。**“胸痛就是心脏病”不是诊断推理，“看到相向就把速度相加”也不是数学推理。只有当至少两个解释都能说明一部分现象，却在某个关键证据上给出不同预测，鉴别才真正开始。

但诊断学也有自己的危险，而且它自己最清楚。候选可以无限增加，检验可以不断推迟。更严重的是那个叫“过早闭合”的毛病：某个解释说通了大部分现象，人就自然停止搜索。最危险的不是没有候选，是某个候选过早获得了免于继续受审的地位。

那么，谁来规定一个候选要承担多少举证责任？证据不足的时候能不能行动？

## 法学：不允许“我觉得这样比较简单”成为理由

先说清楚这里用的是法学的哪一块。

不是法条，是**证据法**——它研究的是：在事实不确定的情况下，一群人怎样通过一套程序作出一个可以执行的判断。

三个概念，对课堂特别有用。

**第一，举证责任。**谁主张，谁举证。这句话的分量在后半段：**如果证据不够，由谁承担不利后果？**这不是修辞问题，是规则问题——民事案件里通常是“优势证据”就够，刑事案件里要求“排除合理怀疑”。**同一份证据，在不同的证明标准下，结论可以完全相反。**

**第二，有效的异议。**在法庭上，“我不同意”什么也不是。有分量的异议必须做到其中之一：指出对方哪一项前提没有证据支持；给出一个能让对方的说法失败的反例；说明补上什么条件对方才能重新成立。

**第三，可撤销的结论。**这一条最反直觉。法律推理不是一条推出来就永远不动的链子——它的结论是“在当前证据和程序下暂时成立”，同时保留上诉、再审、证据排除的通道。

于是有了一件很微妙的事：**一个判决之所以可以拿去执行，不是因为它绝对不会错，而是因为它明确写清了什么新证据能重新开启争议。**

这正是数学课堂缺的那一样。课堂上的“得出结论”往往像终审判决：板书一写，后面的练习只负责套用。**没有人告诉孩子，什么情况下这个结论该被重新打开。**

**法学自己也有一个解决不了的问题：**程序可以在纸面上跑得极精密，主张、证据、反例、裁决一应俱全，却始终没有真正拿这个结论去做一件会暴露新信息的事。**争论越充分，越容易把行动推迟到“完全确定之后”。**

课堂讨论里，孩子常说：“我觉得这样算简单。”“我跟他思路不一样，但都对。”

这些话表达了立场，却没有建立一个**可以裁决的争议**。只要每个人都能保留“自己的想法”，讨论就不会产生结构性的后果——最后通常由老师宣布哪种方法更规范，争论在权威总结里结束。

法学面对的从来不是“让每个人充分表达”，而是**怎样在冲突的主张之间分配证明的风险**。

一方提出主张，不能只说“我相信”。他要承担举证责任：提供什么证据、达到什么标准、证据不足的时候由谁承担不利后果。另一方不只是提出不同观点，还可以质疑证据的可靠性、提供反例或替代解释。裁判者也不能把“我更喜欢这个说法”写进判决，而要说明哪些事实被认定、哪一方没有完成证明责任。

数学课堂从法学拿到的不是“把教室变成法庭”，而是三条更硬的纪律。

**第一，候选方法必须承担举证责任。**

一个孩子说“这题用比例”，他要证明的不只是能列出一个比例式，而是：哪两个量的比在变化前后保持不变；对应项为什么这样排；哪个条件一变这个关系就不成立；跟另一个候选相比，什么证据更支持比例而不是“每次多加固定的量”。

**方法一旦不用承担这些责任，就会变成关键词触发器。**

**第二，反对必须能改变裁决。**

"我不同意"不够，"还有别的方法"也不够。有效的异议要指出：对方哪一项前提没有证据支持；什么反例能让他的方法给出错误预测；补上什么条件他的主张可以重新成立。

比如有孩子说"两个角相等的三角形一定全等"。反对者不能只说"这是错的"，要拿出两个形状相同、大小不同的三角形，说明**两角相等只锁定形状，不锁定尺度**。这个反例改变了主张的状态。

### 第三，结论只拿到临时的地位。

课堂上的"得出结论"往往像终审判决：板书一写，后面的练习只负责套用。法学却始终保留改判的通道。一个裁决之所以可以行动，不是因为它绝对不会错，而是因为它在当前的证据下暂时成立，同时明确了什么新证据能重新开启争议。

但法学也有一个它自己解决不了的问题：对抗程序可以在纸面上跑得非常精密，孩子列主张、列证据、列反例，却始终没有真正用这个结论去做一件会暴露新信息的事。**争论越充分，越可能把行动推迟到"完全确定之后"**。

现实中的导航不能这样。

## 导航学：先想清楚再行动，本身会让人迷路

### 先说清楚导航学在这里研究什么。

不是地图学，是**动物和人在不确定的环境里怎么知道自己在哪儿**。

有两套完全不同的机制，各有各的毛病。

**第一套叫路径积分**。不看外面，只根据自己走过的方向和距离，一路推算相对位置。撒哈拉的沙漠蚁是这方面的高手：它在毫无地标的沙地上曲折觅食一两百米，找到食物之后能**沿直线**直奔巢口。它没有地图，它一路在算。

人也有这套本事——闭上眼在熟悉的房间里走几步，你大致知道自己在哪儿。

它的毛病是**误差会累积**。每一步的小偏差都加进去，走得越远越离谱，而且它自己不知道自己错了。

**第二套是地标导航**。看外面：路口、大楼、一棵特别的树。地标的好处是能**一次性把误差清零**——认出那个红色的加油站，你立刻知道自己在哪儿。

它的毛病是**地标会骗人**。地标会移动、会被遮挡、会长得很像。实验里有个很有意思的结果：给人一个虚拟环境，如果偷偷把地标挪个位置，人会被它带偏；而当地标显得不太可靠时，人会**自动转回去**依赖那个比较粗糙的整体空间感。

所以导航研究真正关心的问题不是"地图重要还是地标重要"，而是：**两套线索冲突的时候，怎么整合、怎么取舍、什么时候切换主导**。

还有一个区分，对课堂直接有用。走偏了，可能是三件完全不同的事：

- **位置错了**——地图对，只是这一步算错了；
- **地图错了**——从一开始就把这片地方认成了另一片；
- **航标错了**——路和地图都没问题，但你事先设的那个"看到什么就说明走对了"的检查点本身选错了，所以它一路都没报警。

**课堂上几乎所有的订正，都默认是第一种。**

导航不是从一张完美的地图开始的。

行动者通常只有部分信息：一个出发点、一个大致方向、几个地标、对自己走了多远的估计，以及不断变化的环境。每走一步，他都要更新"我现在在哪里"。

一种基本机制叫**路径积分**：根据自己走过的方向和距离持续推算相对位置。它不依赖外部地标，但会累积误差。另一类机制依靠**地标**重新定位：地标能纠正内部误差，但地标可能移动、相似、被遮挡，或者本身就不可靠。

导航研究关心的不是“地图重要还是地标重要”，而是不同线索在冲突的时候怎样被整合、竞争和取舍。虚拟环境的实验发现：人在有分散地标的时候更容易发现新捷径，但也容易被移动过的地标带偏；当地标显得不可靠时，人会转而依赖比较粗略的整体空间感。

这给数学学习第三条路：

**不要等到完全确定才使用一个关系。带着一个临时的判断进入行动，让行动产生新的位置证据，在误差累积起来之前重新定向。**

传统的“检查”通常发生在答案之后：重算一遍，代回去看看。导航式的检查必须发生在**途中**，而且检查点不能全部由老师提供。

还是解  $3(x-2)=2x+5$ 。出发之前，孩子可以先布下三个“航标”：

1. 去括号之后，等式两边表示的仍然是同一个未知量关系；
2. 把含  $x$  的项挪到同一边之后， $x$  的系数应该由 3 和 2 的差决定；
3. 结果代回去的时候，不只是两边数值相等，原题里的运算顺序也要被满足。

**这三条不是步骤提示。**步骤提示告诉孩子“下一步移项”；航标告诉他“走到这里的时候你应该看见什么，如果没看见，说明路线或者定位出了问题”。

本章开头那个孩子，写下  $3x-2$  的时候，如果他事先布过第二条航标，他会发现系数对不上——**这个发现由他自己作出，不需要老师在旁边。**

导航学还给了一个特别有用的区分。走偏了，可能是三种完全不同的事：

- **位置错了**：整体的关系地图是对的，只是局部算错了、符号写反了。
- **地图错了**：整个问题表征就选错了，比如把追及看成了相遇。
- **航标错了**：路线和地图都没问题，但事先设的检查点本身就不对，所以它一路都没报警。

三种错，处理方式完全不同。而课堂上几乎所有的“订正”，都默认是第一种。

## 它们共同踩着的那块地

三门学问的词，一张表收好。

| 这个词   | 一句话解释               | 在这一章里对应什么       |
|-------|---------------------|-----------------|
| 问题表征  | 把叙述压缩成有区分力的一句话      | 把题面压成能制造分歧的表征   |
| 鉴别诊断  | 同时挂着两个候选，去找能分开它们的证据 | 至少两个会导致不同下一步的候选 |
| 过早闭合  | 说通了就停止搜索            | 第一个想法太早获得豁免     |
| 举证责任  | 谁主张谁举证，证据不足谁吃亏      | “这题用比例”要证明什么    |
| 证明标准  | 要多硬的证据才算数           | 几个例子够不够         |
| 可撤销结论 | 暂时成立，但写明什么能推翻它      | 可改判的临时裁决        |
| 路径积分  | 靠自己走过的方向和距离推算位置     | 一步步往下算          |
| 地标    | 用外部参照一次性纠偏          | 途中该看到的关系航标      |

|      |               |              |
|------|---------------|--------------|
| 误差累积 | 小偏差越滚越大，而且不报警 | 一路自我感觉良好直到最后 |
|------|---------------|--------------|

三家在打架：诊断学要先把征象压缩成能制造分歧的表征；法学要求主张承担举证责任、结论保留改判通道；导航学则说别等了，先动起来，让行动产生证据。

**但三家共享一个前提：正确的方向，可以在行动之外被确定。**

医生可以在检查前形成鉴别；法庭可以在裁决前完成举证；哪怕导航，也默认存在一个“正确位置”等着被逼近。

学校把这个前提用到了极致：先确定题型，选一个正确方法，执行并检查；错了就归因为“没认准类型”“没记住方法”“执行不够准确”。

**推翻它的材料来自导航学自己：**很多位置知识只有行动之后才出现。船不会等海图、天气和定位全部无误才离港；人在陌生城市不能先证明所有道路可靠再迈步。**你不走，就永远拿不到那一份只有走了才会出现的证据。**

#### 四、五步校航，和那张格子

于是我把这条路线整理成五步。它不是解题步骤，是一条在不确定里前进并且保持定位的路线。

**第一步，诊断：**把题面压缩成一份能够制造分歧的关系表征。

**第二步，举证：**让至少两个候选各自承担证明责任，而且这两个候选必须会导致不同的下一步。

**第三步，起航：**作出一个可以改判的临时裁决，选一条最能产生信息的路走。

**第四步，校偏：**把实际看到的和事先布下的航标对照。

**第五步，重定位：**判断这次要改的是位置、地图，还是航标。

第五步完成之后，回到第一步——**新的问题表征就是下一轮的起点。**

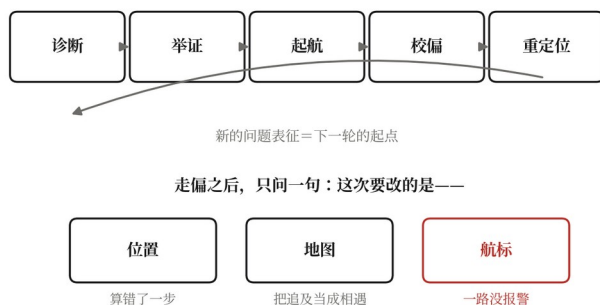


图6 五步校航，与走偏之后的那一问：改位置、改地图，还是改航标？

其中第三步最反直觉，值得单独说。所谓“可以改判的临时裁决”，是让孩子在动笔之前写下四句话：

- 我暂时采用哪个关系或表示；

- 哪一份证据让它领先；
- 哪个替代解释还没有被完全排除；
- **看到什么，我必须重新开启争议。**

最后一句是全书性价比最高的一句话之一。它把"检查"从答案之后挪到了出发之前。

回到那张格子。这一章的竖轴是：**方向是谁给的？**

|         | 这次走对了                        | 这次走偏了                               |
|---------|------------------------------|-------------------------------------|
| 方向自己拿得住 | 活的正确：他知道自己凭什么选这条路，也知道什么时候该掉头 | 正在发生：偏了，但他自己发现了，并且能说出是位置、地图还是航标出了问题 |
| 方向由外面给  | 导航式的正确：一路没错，撤掉提示就不知道在哪儿      | 普通的走错                               |

## 五、六个场面，六次自己定位

(以下场景为构想，题目均为常见题型。)

### 场面一：分数除法不是"乘倒数"，是单位重新定位

"6 里面有几个  $\frac{1}{2}$ ?" 和 " $\frac{1}{2}$  的多少倍是 6?" 在算式上都是  $6 \div (\frac{1}{2})$ ，但它们问的不是同一件事。

**诊断**要先把题压缩成："现在以什么为一份?" 两个候选：以 1 为一份，还是以  $\frac{1}{2}$  为一份。这两个候选会导致不同的下一步。

孩子如果只记住"除以分数等于乘倒数"，他从来没有做过这一次诊断。等到分数除法应用题一来，他就只能靠猜。

### 场面二：把"共同变大"送上证据席

孩子说"这两个量一起变大，所以是正比例"。

按法学的纪律，这是一个需要举证的主张。举证内容是：对应值的比是不是同一个常数；有没有起始量。

反对者的任务是造一个反例：**每千米 2 元，但起步价 5 元**——两个量确实一起变大，比值却不恒定。

这一个反例，比讲十遍定义管用。

### 场面三：增根不是计算事故，是路线改变了解集

分式方程去分母之后出现增根，几乎所有孩子都被教成"最后要检验"。

用导航的说法：**去分母这一步，改变的是地图，不只是位置。**原来的方程带着定义域，去分母之后的方程不带。所以事先就该布一个航标："这一步会不会改变解集?"

布下这个航标的孩子，检验不是老师要求的仪式，是他自己设的报警器。

## 场面四：辅助线不是灵感，是一次要产生证据的行动

孩子常说“我看不出来该怎么添线”。而做得好的孩子，他的辅助线也是试的，只是他试之前先说了三句话：

- 这条线预计制造哪两条新关系；
- 这些关系怎样接近目标；
- **如果第一条预计的关系没有出现，什么时候撤掉这条线。**

第三句是“起航”和“乱试”的分界。乱试的人画满十条线不肯擦；会导航的人两条线之内就知道要退回来。

## 场面五：坐标系本身可以被改判

学函数图像的时候，坐标系通常是给定的。孩子默认它是世界的一部分。

真正的重定位发生在这一刻：“**如果我把原点挪到这里，这个问题会不会变简单？**”

这不是技巧，是一次“改的是地图不是位置”的判断。

## 场面六：概率模型要预先写下会推翻自己的观测

在做任何概率估计之前，先让孩子写一句：“如果出现什么结果，我就该承认这个模型不对。”

写不出来，就说明这个模型还不是一个可以检验的主张——**它只是一个说法。**

## 六、明天就能用的一张卡

**出发之前问三句：**

1. 至少哪两个诊断，会给出不同的下一步？
2. 我暂时选哪一个？什么证据让它领先？
3. **途中应该看到什么？看到什么我必须改判？**

**走偏之后问一句：**

**这次要改的，是位置、是地图，还是航标？**

这一问是这一章最值钱的东西。它把“订正”从“抄一遍正确步骤”变成了一次真正的诊断：算错了一个数，改位置；把追及当成了相遇，改地图；一路自我感觉良好直到最后才发现不对，那就是航标设错了——**下次要换一个检查点。**

**给老师的三条提醒：**

- **别做稻草人。**两个候选里至少要有个来自孩子当前真实的倾向。老师虚构一个明显错误的选项，只会训练孩子猜老师想要哪个。
- **两个候选就够。**不要为了显得“思维开放”而列五种。**候选的标准不是多，是会导致不同的下一步。**
- **偏差出现的时候先别急着纠正。**那一刻正是这一章说的全部东西发生的地方。给他十秒钟，问一句“你觉得刚才哪一步开始不对？”——这十秒钟，比替他改十次都值。

## 关于“等一等”，有一组测过的数字

上面那句“给他十秒钟”，我要按实测数据修正一下——因为这件事早在五十年前就被  
人拿秒表量过了。

上世纪七十年代，有研究者做了一件很笨也很有用的事：录下大量课堂，逐次测量老师提问之后到底等了多久。

结果是：**平均一秒**。提完问题，等不到一秒就会开口——重复一遍、换个说法、点另一个人、或者自己把答案说了。学生答完之后到老师开口，平均**0.9秒**。

然后研究者训练老师把这两个停顿**延长到三到五秒**。变化是成组出现的：

- 学生回答的**长度增加**；
- **答不上来的次数减少**；
- 学生开始给出**带猜测性质的回答**（“我觉得可能是……因为……”）；
- 学生**主动提问**的次数增加；
- **平时被认为反应慢的孩子**，发言次数明显增加。

最后这一条值得单独说一遍。**不是那些孩子突然变聪明了，是这间教室原来的节奏，从来没给过他们开口的机会。**

所以我把话改准一点：**不是十秒，是三秒**。三秒是有实测支撑的门槛，十秒是野心。而这三秒之所以难，正是因为它在教室里长得离谱——**一秒是常态，三秒是常态的三倍**。

这也解释了为什么本章开头那二十秒会那样发生。老师不是不愿意等，是一秒之后那份安静已经变得难以忍受。

## 七、这套眼光的分寸

**第一，不是每道题都值得诊断。**

练习题就是练习题。一节课里，值得走完五步的通常只有一到两道。**其余的题，就让孩子快快地做对**。把每道题都做成鉴别诊断，孩子会累垮，而且会失去应有的熟练度。

**第二，不要把孩子的每一次判断都变成审判。**

举证、反驳、裁决这套说法是给老师看的结构，不是给孩子的课堂用语。**没有哪个十岁的孩子需要听见“你没有完成举证责任”**。落到课堂上，它就是一句话：“你凭什么这么想？还有别的可能吗？”

**第三，低年级用两个候选，不用五步。**

完整的五步适合小学高年级和初中。低年级只做两件事就够：**先猜一下会是多少，再看看和你猜的差多远**。这已经是完整的校航循环的最小版本。

**第四，“不给答案”远远不够。**

现在很多人用 AI 辅导时会设定“不要直接给答案，要引导”。这解决不了这一章说的**问题——引导本身就是方向性线索**。孩子仍然没有机会自己形成预测、自己发现偏差。

真正要改的是顺序：**让孩子先写下他的临时裁决和改判条件，再让 AI 说话**。这一个顺序上的改变，比任何“引导话术”都管用。

**第五，“可以改判”不是让人怀疑一切。**

数学证明的确定性不受影响。这一章说的是学习过程中对方法和问题表征的暂时选择；一旦证明完成，结论就是结论。**要求孩子对已经证明的东西继续保持怀疑，那是另一种毛病**。

## 这一章的一句话

走对每一步的人，不一定知道自己在哪儿。真正的方向感，是在第一次走偏的时候长出来的——前提是，那一次没有人抢在他前面把方向盘转过去。

下一章讲第三条路，也是最难的一条：孩子已经很会用一个方法了——**难的是让他对“几乎一样”的那道题，说不。**

## 第六章·对"几乎一样"说不

第二编 一堂课该留下什么·之三

### 一、会做十种变式的孩子，为什么可能更危险

下面这个场面是我构想的，题目本身极常见。

六年级，一个孩子正比例学得很好。老师给过他十几种变式：路程和时间、总价和数量、面积和边长、工作量和工作时间、图上距离和实际距离。他每一种都做得又快又准。现在给他这道题：

**某种出租车，起步价 5 元（含 2 千米），之后每千米 2 元。行驶 10 千米付多少钱？行驶 20 千米呢？20 千米的车费是不是 10 千米的两倍？**

他看了一眼，说："这是正比例。千米数变成两倍，钱就变成两倍。"

这不是他"不认真"。恰恰相反——**是他太会认了。**

"每千米 2 元"这五个字，在他学过的十几种变式里出现过太多次。它太像正比例了。像到他的判断在看清楚道题之前就已经完成了。

而那个只做过三道正比例题、还有点不熟的同学，反而更可能停下来数一数：走 2 千米是 5 元，走 4 千米是 9 元，5 和 9 不成倍数——**因为他还不熟练，所以他还需要去看。**

我要说的不是"熟练有害"。熟练当然好。

我要说的是：**一个熟练的方法，如果没有配上一条同样清楚的"什么时候不许用它"，它就会自己往外扩张。**而且扩张的时候没有任何报警——因为它扩张到的那些题上，它看起来完全合理。

我把这一格叫作**过界的熟练**。

### 二、这一章要还的债

一个概念真正的边界，不只由"哪些例子该用它"构成，也由"哪些几乎一样的例子必须忍住不用"共同构成。

我们的教学几乎只做前半边。正例讲得很足，变式给得很多，反例通常只在最后作为"注意事项"出现一下。**而边界恰恰长在后半边。**

### 三、三门学问，三种"还算不算同一个"

**音乐学：所有的音都变了，还是同一首歌**

**先说清楚音乐学在这里研究什么。**

不是好不好听，是人凭什么**认出一首曲子**。

先做个实验。你会唱《生日快乐》。现在请一个人用比平常高很多的调唱它——每一个音的绝对高度都变了，一个都不剩。你还是一听就认出来。

这件事其实相当不寻常。**你认出的东西，一个音符都没有留下。**

那你认的是什么？是**音与音之间的距离**。音乐里管两个音之间的距离叫**音程**。《生日快乐》开头是"同一个音重复一次，然后往上跳一点，再回来"——只要这一串相对距离保持着，整体搬到哪个高度都还是它。把整首曲子平移到另一个高度，音乐上叫**移调**。

这不是玄学，是可以测的。心理学实验里给受试者听一段短旋律，隔一会儿再听一段移调过的，让他判断是不是同一段——受过训练的人能相当准确地抽取音与音之间的关系来判断，而不是记住某个具体的音高。

变化还能更大。改节奏、加装饰音、换乐器——只要某些承重的关系还在，听的人仍然会说“这是那个主题”。作曲家整本整本的变奏曲，靠的就是这件事。

**但音乐同时给了一个警告，而且实验里是专门设计出来的。**研究者会故意加进一种**诱饵旋律**：它保留了原曲的大致起伏轮廓，却改掉了几个关键的音程。结果是——**很多人会把它误认成原曲。**

**相似会制造误认。**音乐可以容忍这种模糊，因为作品的生命有时恰恰来自暧昧、引用和变形。

**数学不能这么宽容。**把“两个量同时增加”当成正比例的轮廓，把“每增加一单位就增加固定的量”当成内部音程——听上去很近，数学后果完全不同。

把《生日快乐》整体升高一个大二度，多数人仍然认得出来。

每一个绝对音高都变了，旋律却没有变成另一首歌。我们认出的不是一组固定的频率，而是**音与音之间的相对距离、起伏的轮廓、节奏的组织 and 重音的位置。**

移调把整组音搬到新的高度，内部关系保持不变；变奏还能改节奏、加装饰音，只要某些承重的关系还在，听的人仍然会说：“这是那个主题。”

音乐心理学对移调旋律识别的研究反复表明：听者能在音高整体改变之后靠关系表征认出旋律。受过训练的听者会抽取相邻和非相邻音之间的音程关系——**移调耐受不是记住了单个的音，而是编码了关系模式。**

这给数学一条很强的路线：**先让一个关系以具体形式出现，再通过系统的“变调”逼学习者从绝对材料转向相对组织，最后在新形式里重新听见同一个主题。**

一道“每盒6个、4盒有多少个”的乘法题，可以变调成点阵、长方形面积、速度与时间、单价与数量；一个分数可以在饼图、数轴、比、除法和算子之间变调；一次函数可以在表格、图像、解析式和真实过程之间变调。

**孩子如果只能在原调上认出关系，他学到的是表面的音高；换了调还认得出，才可能抓住了关系结构。**

**但音乐的宽容带来一个危险。**旋律的身份并不是无限宽的。只保留大致的起伏而改掉关键音程，听者可能把一段**诱饵旋律**误认成原主题。心理学实验里专门会加入“保留轮廓的诱饵”，正因为相似会制造误认。

音乐可以容忍这种模糊——作品的生命有时恰恰来自暧昧、引用和变形。

**数学不能这么宽容。**把“两个量同时增加”当成正比例的轮廓，把“每增加一单位就增加固定的量”当成内部音程，两者听上去很近，数学后果完全不同。

音乐教会我们把“同一个”的范围扩大，却没给出一道足够严格的边界。

## 密码学：相似不是证据

**先说清楚密码学在这里的两个基本装置。**

密码学不是“加密”两个字那么简单。它有很大一块处理的是另一个问题：**我怎么确认这份东西没被人改过、而且确实是他发的？**

**第一个装置：摘要。**

有一类算法（哈希算法）能把任意长度的内容，压成一小串固定长度的字符——比如一整本书压成64个字符。这串东西叫**摘要**。

它有个要命的性质：**内容哪怕只改一个标点，摘要就会完全变样，而且是面目全非，不是“差一点”。**所以只要对比摘要，就知道这份东西还是不是原来那一份。

## 第二个装置：签名。

签名者手里有一把只有他有的**私钥**，用它给内容生成一个签名；任何人拿他公开的**公钥**都可以验这个签名。验通过说明两件事：内容没被改，而且确实出自他。

这两个装置背后是同一种冷酷的态度：

**相似不是证据，熟悉不是证据，说得很像也不是证据。要么给出可以复核的对应关系，要么就不算。**

这和音乐正好顶上了。音乐里的合法变调可以让表面材料大改而仍算同一个；密码学要求：**变了就得重新签，旧签名不能沿用。**

还有一个装置顺带提一下，因为它对老师有用。有一类叫**零知识证明**的方法，能做到一件听起来不可能的事：让对方相信“我知道这个答案”，却**不透露答案本身**，也不透露我是怎么知道的。

它对课堂的提示是：**我们不必读取孩子每一秒的内心活动，才能确认他学会了。**他提交一份足以验证关系身份的**证据就够**——不必被要求公开所有的草稿、犹豫和私人联想。

一份电子文件发给别人，接收者没法靠“内容看起来差不多”判断它是不是完整的，也没法只凭署名确认是谁发的。

密码学的态度非常冷酷：**相似不是证据，熟悉不是证据，说得很像也不是证据。**

哈希把任意长度的信息压缩成固定长度的摘要，用来检测信息在生成摘要之后有没有被改动。数字签名让签名者用私钥生成签名，接收者用对应的公钥验证数据有没有被修改、签名来自谁。**一份消息哪怕只改了一个字符，也不能因为“意思基本没变”而沿用原来的签名。**

这跟音乐形成尖锐的对立：音乐里的合法变调可以让表面材料大幅改变而仍算同一个；密码学要求验证程序在明确的规则下给出接受或者拒绝。

它给数学学习的路径是：**每一次声称两道题属于同一个关系，都要生成一份别人能独立复核的“关系签名”。**

关系签名不是知识点的名称。孩子说“这是正比例”，没有完成签名；说“两个量一起增加”，更没有。一份最低限度的签名要包含五项：

1. 哪些对象在被比较；
2. 以什么为单位；
3. 哪个关系在允许的变化中保持；
4. **哪个承重条件一旦改变，原关系就失效；**
5. 用什么操作或观测，第三方可以复核。

比如正比例的签名不能只写“一个量变大另一个也变大”，要写：**在同一计量范围内，对应量的比值保持同一个常数，图像过原点；如果存在固定的起始量，比值就不再恒定。**

相似三角形的签名不能只写“形状一样”，要说明对应角相等、对应边成比例；要全等，还得比例因子为1。

等式变形的签名不能只写“两边做同样的操作”，要声明这个操作在当前定义域里是否可逆——平方、乘以含未知数的式子、约去可能为零的因子，都可能只保留单向推出。

**密码学还有一个提醒对老师很有用：**验证不等于泄露全部生成过程。我们不必读取孩子每一秒的内心活动才能确认他学会了。**他提交一份足以验证关系身份的证据就够了，不必被要求公开所有的草稿、犹豫和私人联想。**

**但密码学也有盲点。**签名验证必须在既定的协议里运行。它能回答"这份数据和签名匹不匹配"，却不回答"即使匹配，现在是不是该行动"。

在课堂上，孩子甚至可能背下一份极其精确的适用条件清单，做题时逐项核对，却仍然没有形成活的边界——**他验证的是老师预先给定的签名，不是他自己在冲突中生成并改写的签名。**

## 免疫学：识别正确，也可能反应错误

**先说清楚免疫学在这里研究的是哪个难题。**

我们通常以为免疫系统的任务是"认出坏东西然后消灭它"。真实的难题要拧巴得多。

免疫系统每天接触的东西里，绝大多数**不能碰**：你自己的细胞、你吃进去的食物、肠道里几十万亿个共生细菌、空气里的花粉和灰尘。而其中很多，**跟真正的病原长得很像。**

更麻烦的是，免疫细胞的受体有**交叉反应**——同一个受体常常能以不同强度识别好几种分子。它做不到一把钥匙只开一把锁。

于是两边都是灾难：

- 对一切"看起来像危险"的东西都强烈反应 → **过敏、自身免疫病**，系统开始打自己；
- 为了不误伤而一律沉默 → 感染和异常细胞长驱直入。

所以免疫学习**同时包含两种相反的工作**：一边形成能识别危险的反应谱，一边**主动删掉和压住**那些会攻击自身和无害对象的反应。

后半有个专门的名字，叫**耐受**，分两道关：

- **中央耐受**：免疫细胞在发育的时候要过一道筛——对自身反应太强的，直接被淘汰掉；
- **外周耐受**：漏网的那些，在身体各处还有一套机制继续压着它们，让它们不反应或者被调节。

**最关键的一句在这里：不响应不是"没认出来"，而可能是认出来之后，被一整套机制主动拦住了。**

而且这个拦截不是简单的开关。免疫细胞会用**结合的持续时间**和一连串的信号步骤来区分：这是短暂蹭到的自身分子，还是一个真该反应的信号。**改掉其中一个校验环节，辨别就崩了。**

一句话记住：**边界的一半是"该认什么"，另一半是"认出来了也得忍住"。**

免疫系统面对的不是整齐分好类的练习册。

它每天接触大量自身分子、食物成分、共生微生物和环境物质，很多对象彼此高度相似，而受体还有交叉反应——同一个受体可能以不同程度识别多个分子。

如果对一切"看起来像危险"的东西都强烈反应，结果不是更安全，而是过敏、自身免疫和组织损伤；如果为了避免误伤而一律沉默，感染又会逃逸。

所以免疫学习不是不断扩大识别范围，它同时包含两种相反的工作：**形成能识别危险对象的反应谱；删除、抑制或调节那些会错误攻击自身和无害对象的反应。**

发育中的淋巴细胞会经历选择，自身反应性太强的克隆可能被删除；逃过这一关的，还有外周的机制通过无反应和调节继续控制。近年的研究更进一步显示：免疫识别不是静态的"钥匙配锁"，细胞会利用结合的持续时间和连续的信号步骤，来区分短暂的自体相互作用和足以启动反应的信号——**改掉其中一个校验环节，配体辨别就会被破坏。**

这给数学一个常被忽略的判断：

**不响应，不是知识的缺席，而是一种学出来的积极结果。**

我们通常把正例当作知识内容，把反例当作检查工具。免疫学要求把这个次序换过来。免疫耐受不是“没认出来”，而可能是在认出相关信号之后，仍然通过阈值、时间和环境的调节，阻止了破坏性的反应。

数学里的“耐受”也不是让孩子忽略错误。它指的是：**当题目和熟悉的模型高度相似时，孩子能够抑制自动调用，保持一个足够长的检查窗口，直到承重条件被确认。**

比如：

- 看见“每千米 2 元”，仍能暂时不启动正比例，先检查有没有起步价；
- 看见“不能同时发生”，仍能暂时不套独立事件公式，先分清互斥和独立；
- 看见“两边平方”，仍能暂时不宣布等价，先检查能不能逆推回去；
- 看见“角分别相等”，仍能暂时不宣布全等，先检查尺度关系。

**关键不是性格谨慎，而是形成了一个说得出口位置的“不触发结构”：**因为哪一个条件还没出现，所以我现在不调用这个关系。

**免疫学还反过来限制了“多做变式”的热情。**如果训练只不断提供同类的正例，识别器会越来越容易被激活；如果正例和错误的近邻在表面上从来没接近过，孩子学到的是粗糙的分界；**如果老师每次都提前标注“这是一道反例”，孩子根本不需要自己抑制，耐受边界就永远长不出来。**

**它们打架的地方，和共同踩着的那块地**

**三门学问的词，一张表收好。**

| 这个词     | 一句话解释             | 在这一章里对应什么      |
|---------|-------------------|----------------|
| 移调      | 整体搬到另一个高度，关系不变    | 合法的变调：换情境、换表示  |
| 音程      | 两个音之间的距离          | 承重的那条关系        |
| 诱饵旋律    | 保留轮廓、改掉关键音程的假货    | 欺骗性近邻          |
| 摘要      | 内容改一个字，它就完全变样     | 承重条件一变，关系就不是它了 |
| 数字签名    | 可被第三方独立复核的对应关系    | 关系签名的五项        |
| 零知识证明   | 让人相信你知道，却不用交出全部过程 | 不必读取孩子每一秒的内心   |
| 交叉反应    | 同一个受体会认上好几种相似分子   | 熟练的方法会往外扩张     |
| 耐受      | 认出来了也主动不反应        | 有证据地不触发        |
| 中央／外周耐受 | 发育时先筛一遍，之后还有第二道   | 边界要在两个地方分别长出来  |

把三家并排写，会得到三句都对的废话：像音乐家一样在变化中抓住不变，像密码学家一样严格验证，像免疫系统一样区分对错。

它们真正的冲突是这样的：

- **音乐学要求扩大同一性**：一个主题要有生命，就不能被锁死在第一次出现的音高上。
- **密码学要求收紧同一性**：合法的改变如果没有重新生成签名，旧签名就不该继续通过。
- **免疫学要求延迟同一性的后果**：即使识别成立，也不能直接把"识别"结算成"响应"。

选音乐学，容易培养宽泛的迁移，却放大误触发；选密码学，容易培养条件核对，却把判断变成执行既定协议；选免疫学，容易培养谨慎的边界，却让孩子因为怕用错而失去生成能力。

**而三家共享着同一个想法：一个概念总能靠某个正面标准来结算。**

音乐靠关系模式，密码靠验证协议，免疫靠反应谱。学校最直接——靠定义。我们默认：只要把定义讲清楚、把正例给够，边界自然就有了。

**推翻它的材料来自免疫学自己**：一个识别系统的边界，有一半是由它学会不响应什么构成的。删掉那一半，剩下的不是"边界更宽"，是**根本没有边界**——只有一个越用越顺、越顺越往外扩的触发器。

## 四、六步校界，和那张格子

于是这一章的路线是六步。它比前两章的循环长一点，因为最后两步是我们平时几乎不做的。

**第一步，取核**：从一个具体解法里取出可以被变调的关系母题——写清对象、单位、关系，以及哪一项撤走关系就变了。

**第二步，变调**：在保持母题的前提下，主动把表面改到足够远。换数字不算，要换情境、换表示、换载体。

**第三步，验签**：为"为什么这仍然是同一个关系"生成一份第三方能打回的签名。

**第四步，近邻**：制造一个**最像正例的错误对象**——表面尽量相似，只破坏一个承重点，而且错误调用会产生看得见的后果。

**第五步，耐受**：面对这个近邻，练习有证据地不触发。

**第六步，改签**：重写签名里最少必要的字段，并且由孩子自己生成下一轮的一正一负。

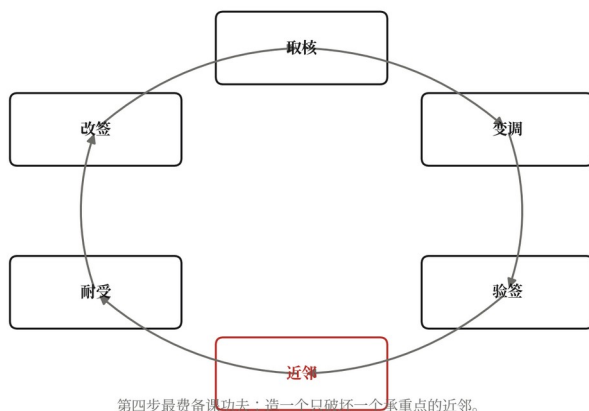


图7 六步校界。第四步最费备课功夫。

第四步是全书最需要备课功夫的一步，也是最值钱的一步。它不是"随便找个反例"—

只破坏一个承重点。破坏两个，孩子分不清是哪一个让它失效的；一个都不破坏，那只是变式。

第五步要的是三句话：我最先想调用的是什么？签名里的哪一项没有通过？所以我在改用什么，或者暂时还不能决定什么？

最后那半句“暂时还不能决定什么”，是数学课上最少听见、也最该听见的话。

回到那张格子。这一章的竖轴是：该不用的时候，忍得住吗？

|          | 该用的时候用得上             | 该用的时候想不起来             |
|----------|----------------------|-----------------------|
| 该忍的时候忍得住 | 活的正确：又用得开，又收得住       | 正在发生：还不熟，但已经知道要先看承重条件 |
| 该忍的时候忍不住 | 过界的熟练：什么都能套上去，而且套得很顺 | 普通的不会                 |

请注意左下角这一格的位置：它在“用得上”这一列。也就是说，从成绩上看，过界的熟练往往是班上最好的那批孩子。这是这一章最需要说清楚的一件事。

## 五、六个场面，六对“最像的错误”

每一个都是一正一负：一个是合法的变调，一个是只破坏了一个承重点的近邻。

先说一个可以省事的地方：诊断题平台 Eedi 公开的题库里，每一个错误选项都事先对应到一个具体的误解——已经整理成册的误解有 2587 条，其中 1604 条在题目中被真正用到过。这是一座现成的“欺骗性近邻”库，不必每一道都自己造。它的详细读数在第七章。

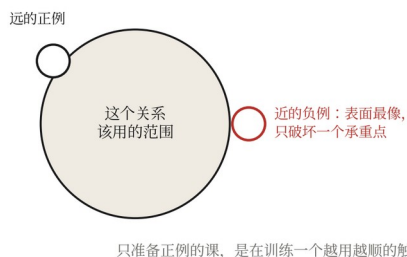


图 8 远的正例负责让关系脱离材料，近的近例负责画出边界。

(以下题目均为常见内容。)

### 场面一：正比例 —— 承重点是“起始量为零”

**合法变调：**路程与时间、总价与数量、图上距离与实际距离——载体全变了，比值恒定、过原点没变。

**欺骗性近邻：**起步价 5 元、每千米 2 元。两个量确实一起变大，只破坏了“起始量为零”这一点。

## 场面二：平均数 —— 承重点是"每个数据同样有分量"

**合法变调：**算平均分、平均身高、平均速度里的一部分情形。

**欺骗性近邻：**一个班 40 人平均分 80，另一个班 10 人平均分 90，两个班合起来的平均分不是 85。人数不同，每个数据的分量就不同了。

## 场面三：方程变形 —— 承重点是"保持等式"和"保持解集"不是一回事

**合法变调：**两边同加、同减、同乘一个非零常数。

**欺骗性近邻：**两边同乘  $(x-1)$ 。等式还是等式，解集变了。

这一对是初中最重要的一对，值得花整节课。

## 场面四：相似与全等 —— 承重点是尺度因子

**合法变调：**旋转、翻折、放到不同位置——形状关系不变。

**欺骗性近邻：**两个角相等的两个三角形。它们确实"看起来一样"，只差一个尺度因子。

## 场面五：函数图像 —— 承重点是"局部像"不等于"整体是"

**合法变调：**同一个线性关系，画在不同的刻度上、不同的区间上。

**欺骗性近邻：**一段曲线在很小的区间里看上去笔直。只破坏了"整体"这一点。

## 场面六：互斥与独立 —— 越像越容易混

**合法变调：**抛硬币、抽卡片、掷骰子——独立性不变。

**欺骗性近邻：**"这两件事不会同时发生"。听起来像"互不影响"，其实恰恰相反：**互斥的两件事，一件发生了另一件就一定不发生，这是最强的相互影响。**

这一对我建议每个初中老师都准备好，因为它是所有概率误用的源头。

## 六、明天就能用的一张卡

上课前，为这节课的核心关系准备一对：

一个远的正例，一个近的负例。

远的正例负责让关系脱离原来的材料；近的负例负责画边界。只准备正例的课，是在训练一个越用越顺的触发器。

课堂上，在孩子动手之前问三句：

1. 你最先想调用的是什么？
2. 签名里的哪一项没有通过？
3. 所以现在改用什么，或者暂时还不能决定什么？

第一问要在他动笔之前问——问晚了，他已经在为自己的答案辩护了。

评价的时候记一件事：

他有没有拒绝过一道题？

一个从来没有对任何一道题说过"这里不能这么用"的孩子，无论分数多高，他的边界还没有长出来。

最后，把出题权交出去一次。

学完一个关系，让孩子自己生成一正一负，交给同桌盲测。**能造出一个“最像正例的错误对象”的孩子，一定已经知道承重点在哪儿了**——因为他必须精确地只破坏那一个点。

这是我知道的、最省力的检验方式。

## 七、这套眼光的分寸

**第一，这一章的比喻不能用在人身上。**

我用了免疫学，但“免疫”“耐受”“排异”这些词绝对不能用来对孩子说。**没有哪个孩子是“过敏型”或者“免疫力低下”**。这一章说的从头到尾是一个关系在一次任务里的触发状态，不是一个人的性格。

一旦有人开始说“这个孩子容易误触发”，这套眼光就变成了一张新的标签纸——那是我最不愿意看见的用法。

**第二，先有熟练，才谈得上耐受。**

一个还没把正比例用顺的孩子，你给他起步价那道题，他不会形成边界，只会更糊涂。**耐受的前提是那个反应本来就足够强**。顺序是：先足量的正例，再远的变调，最后才是近邻。

**第三，不要制造混乱型的陷阱题。**

欺骗性近邻必须在数学上干净：**只破坏一个承重点，而且这个破坏是可核验的**。靠歧义的措辞、生僻的知识、故意的漏字来“考倒”学生，那不是近邻，那是刁难。两者的区别在于：**近邻被识破之后，孩子学到的是一条边界；刁难被识破之后，孩子学到的是“要小心老师”**。

**第四，无限的变式不等于更好的边界。**

现在生成十道变式题只要几秒钟。但如果这十道题全是正例，它们只是把触发器练得更灵敏。**变式的价值在于“远”，不在于“多”**。一道远的正例加一道近的负例，胜过二十道同调的练习。

**第五，“暂时不能决定”要被当成正确答案的一种。**

如果孩子说“这道题我还不能判断，因为不知道有没有起始量”，而我们给他打个叉，那这一整章就白讲了。**这句话应该得分**。它是这一章想要的那个结果本身。

## 这一章的一句话

一个方法真正被学会的标志，不是他能用它做多少题，而是他第一次对一道很像的题说：**这里不能用它——并且说得为什么**。

第二编到这里结束。三章讲了三条路：**让这节课留下一节绕不开的东西（第四章）、让方向在走偏时由他自己重新拿住（第五章）、让边界在“几乎一样”处长出来（第六章）**。

三条路合起来，才是“一堂课该留下什么”的完整回答。

而接下来的第三编，要问一个不太好受的问题：**这些事，为什么在真实的课堂里这么少发生？**我会给出三个答案，而这三个答案有一个共同点——**它们说的都不是有人做错了什么**。

## 第三编 为什么它这么少发生

第七至九章

## 第七章·正确来得太早

第三编 为什么它这么少发生·之一

### 一、同一个答案背后，住着四个不同的孩子

下面这个场面是我构想的，但它每天都在一年级教室里发生。

题目： $13 - 8 = ?$

全班四十个孩子，三十六个写下5。老师看了一眼作业，很满意：这个知识点掌握得不错。

现在我们把这三十六个5拆开看。至少有四种完全不同的东西藏在里面：

**第一个孩子**知道13是1个十和3个一，他把那个十拆成10个一，用10减8得2，再加上3，得到5。

**第二个孩子**用“想加算减”：因为8加5等于13，所以13减8等于5。

**第三个孩子**什么也没想，他背熟了 $13 - 8 = 5$ 。

**第四个孩子**用的是“大数减小数”：他看见8和3，算 $8 - 3 = 5$ 。

**四种规则，同一个答案。**

前两个孩子建立了可以生长的关系。第三个孩子有一张表。第四个孩子的规则是错的——而这道题恰好让他的错误规则给出了正确答案。

如果接下来的练习是 $12 - 7$ 、 $14 - 9$ 、 $15 - 8$ ，第四个孩子会一直对下去。他的错误要等到 $13 - 5$ 才会暴露：那时候他会写2，而正确答案是8。

问题来了：在他做对的那几十道题里，我们有任何机会发现他吗？

**没有。因为那些题，正确的规则和错误的规则给出的答案是一样的。**

这就是这一章要讲的事情的起点。它不是“孩子不认真”，也不是“老师不负责”。**是这些题本身，分辨不出他们。**

### 二、这一章要还的债

数学学习很少发生，不是因为错误太多，不是因为反馈太少，也不是因为孩子理解力不够，而是因为课堂过早地生产了正确。

这种正确，出现在三件事发生之前：不同的内部规则被区分开来之前；预测和结果之间的差异被孩子自己拥有之前；私下的疑问进入公共空间之前。

它在这三件事发生之前，就把问题终结了。

### 三、三门学问，三种“正确来得太早”

**统计学：同一个数据，可能出自不同的模型**

先说清楚统计学在这里问的是什么。

这个概念叫**可识别性**。它问的不是“数据够不够多”，而是一个更狠的问题：

**就算给我无限多的这种数据，我能不能唯一地确定，到底是哪个模型产生了它？**

举一个不用公式就能懂的例子。

假设你只能看到两个数的和。有人告诉你和是 10。那么原来的两个数是 4 和 6，还是 3 和 7，还是 1 和 9？

你说不出来。而且再给你一万个“和是 10”的观测，你还是说不出来。这不是样本量的问题——是“只看和”这个观测方式本身，把这些可能性压成了同一个数。

统计学管这种情况叫“不可识别”。解决办法不是收集更多同类数据，而是换一种观测：只要再告诉你它们的差，问题立刻就解开了。

所以统计学真正的功夫在设计上：去找那种能让两个候选模型给出不同结果的观测。

把这件事搬进课堂，那道  $13 - 8$  就是一个“只能看到和”的观测。四个孩子四种规则，产出同一个数。再做二十道同类题，还是同一个数。

于是我们得到一条可以直接改变明天备课的判断：

**一道题，如果目标规则和常见的错误规则会给出相同的答案，它只能用来练习，不能用来诊断。**

**统计学也有它管不到的地方：**它能告诉你什么样的观测有区分力，却不能告诉你孩子脑子里到底有没有一个“模型”。它给的是设计任务的原则，不是心理的事实。

统计学里有一个基本概念叫**可识别性**。它问的是：给定我能观察到的数据，我究竟能不能唯一地确定产生这些数据的模型？

如果两个不同的模型对同一批数据给出完全相同的预测，那么无论收集多少这样的数据，都无法把它们分开。**这不是样本量不够的问题，加多少数据都没用——是这个观测设计本身没有区分力。**

统计学的做法是改变观测设计：换一个条件、加一次干预、换一个测量角度，让两个模型给出不同的预测。

把这件事搬到课堂，那道  $13 - 8$  就是一个不可识别的观测。四个孩子四种规则，产出同一个数。再做二十道同类题，仍然是同一个数。

于是我们得到第一条判断，它可以直接改变明天的备课：

**一道题，如果目标规则和常见的错误规则会给出相同的答案，它只能用来练习，不能用来诊断。**

这句话值得抄下来。绝大多数课堂练习都是练习题，而我们一直把它们当诊断用。

**但统计学也有它管不到的地方：**它能告诉你什么样的观测有区分力，却不能告诉你孩子脑子里到底有没有一个“模型”。它给的是设计任务的原则，不是心理的事实。

## 神经科学：没有属于他自己的意外，路径不会改道

**先说清楚神经科学在这里的那个发现。**

上世纪九十年代，有一组关于猴子的实验，几乎重塑了我们对“学习何时发生”的理解。研究者记录猴子脑内一类多巴胺神经元的放电，同时给它果汁。

- **一开始：**果汁一到，这些神经元剧烈放电——像是在说“有好事”。
- **接着：**先亮一盏灯，过两秒再给果汁，反复多次。慢慢地，放电从果汁移到了灯上。灯一亮就放电，果汁真的来了，反而没什么动静。
- **最有意思的第三步：**灯亮了，但果汁没来。这时候放电跌到基线以下——像是在说“说好的呢”。

结论是：这些神经元编码的**不是奖励本身，而是奖励和预期之间的差**。学界管它叫**预测误差**。

这条机制对我们的用处，全在“预期”这两个字上：**必须先有一个预期，才可能有落差；没有落差，系统就不更新。**

一个孩子如果从来没有形成过预期，那么正确答案出现的时候，他那里什么也不会发生——**不是因为他不用心，是因为没有可供比较的东西。**

课堂上最常见的一幕：孩子看着题，停了三秒。老师看见他停住了，马上说“想想我们昨天讲的那个方法”。这句话极其善意，而且往往有效。

**但那三秒钟，正是他要形成预期的时刻。提示抢在了预期前面。**

这里要区分三种老师的话，区别很硬：

- **情绪支持**（“别急，慢慢来”）——不影响；
- **任务澄清**（“这句话的意思是每人分到同样多”）——通常不影响；
- **方向性线索**（“想想昨天那个方法”“注意括号”）——**这一种会顶掉预期。**

前两种随时可以给。第三种要看时机。

学习为什么会发生？有一条被反复验证的机制是：**当结果和预期不一样的时候，系统才更新。**

关键在于“预期”这两个字。**必须先有一个预期，才可能有落差。**

一个孩子如果从来没有形成过预期，那么正确答案出现的时候，他那里什么也不会发生——不是因为他不用心，而是**没有可供比较的东西。**

课堂上最常见的一幕是这样的：孩子看着题，停了三秒。老师看见他停住了，马上说：“想想我们昨天讲的那个方法。”

这句话极其善意，而且往往有效——孩子立刻就做出来了。

但那三秒钟里，正是他要形成预期的时刻。**提示抢在了预期前面。**后面即使出现了正确答案，那个落差也无从产生，因为从来没有一个属于他的预期在那里等着被否定。

于是我们得到第二条判断：

**要判断学习有没有发生，得看方向性的线索出现在承诺之前，还是之后。**

请注意“方向性”三个字。这里要区分三种老师的话：

- **情绪支持**（“别急，慢慢来”）——不影响。
- **任务澄清**（“这句话的意思是每人分到同样多”）——通常不影响。
- **方向性线索**（“想想昨天那个方法”“注意括号”）——**这一种会顶掉预期。**

前两种可以随时给。第三种要看时机。

**时机有多窄？**实测的数字是：老师提问之后平均等一秒就会开口。也就是说，那个“方向性线索抢在预期前面”的窗口，通常只有一秒宽。（第五章有这组数据的全貌。）

## 信息经济学：全班都对了，班级就失去了独立信息

先说清楚信息经济学在这里的那个模型。

它叫**信息级联**，用一个吃饭的场景就能讲明白。

你到一条陌生的街上找馆子。两家店，A店门口排着队，B店空着。你自己其实更想吃B，但你想：那么多人排A，他们大概知道点什么。于是你排了A。

下一个人走过来，看到A店排队的人更多了——他也排A。再下一个也是。

**问题在哪儿？**你排 A 的时候，把前面那些人的选择当成了很多份独立的证据。可实际上，他们中的大多数，也只是在跟随第一个人。一份信息被当成了几十份。

而且更糟的是：**从你选择的那一刻起，你自己的判断就退出了公共视野。**后面的人看见的只是“又多了一个排 A 的”，看不见你其实想吃 B。

**级联建立的同时，信息就停止流动了。**

经济学家在实验室里把这件事做成了可控实验：让人依次猜一个袋子是哪一种，每个人都能私下看到一点线索，也能看到前面所有人的选择。结果很稳定——**一旦前面两个人的选择一致，后面的人就会开始无视自己手里的线索。**

课堂上的级联比市场上更快，因为它有一个天然的“第一个人”。老师问“这道题选哪个”，成绩最好的那个孩子举手答了，接下来全班的手都举起来。

从表面看，这是一个理解程度极高的班：意见高度一致，正确率接近百分之百。

从信息的角度看，这个班此刻**只有一份判断。**

其余三十九份没有消失，它们只是退回了私下——**不是被隐瞒，是被级联挤出了公共空间。**

第三门学问看的不是一个孩子，是一群人。

信息经济学里有一个模型叫**信息级联**：一群人依次作决定，每个人都能看见前面人的选择。理性的做法是把别人的选择当作信息——如果前三个人都选了 A，即使我自己的判断偏向 B，我也可能跟着选 A。

问题在于：**第四个人跟随的时候，他把前三个人的选择当成了三份独立的证据。而实际上，第二个人和第三个人可能也只是在跟随第一个人。**

一份信息被当成了三份。而且从这一刻起，后面所有人的私人判断都不再进入公共空间——**级联建立的同时，信息就停止流动了。**

课堂上的级联比市场上更快。老师问“这道题选哪个”，班上成绩最好的那个孩子举手答了，接下来全班的手都举起来了。

从表面上看，这是一个理解程度极高的班级：意见高度一致，正确率接近百分之百。

从信息的角度看，这个班级此刻**只有一份判断。**

其余三十九份判断没有消失，它们只是退回了私下。而老师从此看不见它们了——**不是被隐瞒，是被级联挤出了公共空间。**

于是第三条判断：

**公开的一致，不等于多份独立的证据。**

**这三门学问不是三个角度，它们互相拆台。**

统计学会说：只要设计出有区分力的任务，模型就能分开。神经科学反驳：任务再好，如果提示抢在预测前面，落差不属于学习者，内部路径照样不更新。信息经济学又反驳这两家：即使任务有区分力、孩子也形成了预测，只要答案先被公开，级联会把所有差异压平——**每个人心里的那个不同，永远不会说出口。**

**一次真算：这些题里，有多少分不出人**

上面这些都是推理。所以我去算了一次。

英国有一个叫 Eedi 的诊断题平台，它的题有一个特点：**每一个错误选项都事先对应到一个具体的误解**——不是“选错了”，是“因为把加法当成先于乘法所以选了这个”。2024 年，

一批带这种标注的题被公开发布，我用的就是这一份：**1869 道数学题，4370 个已经标注了误解的错误选项，涉及 1604 种不同的误解，分布在 757 个知识点上。**

（下面每一个数字都可以复算。这一节是全书唯一一处用真实数据说话的地方，所以我把口径写清楚。）

**第一个数字：一道题装不下一个知识点上的误解。**

同一个知识点（比如“计算一个数的平方”）在这份题库里平均记录了 **4.5 种** 不同的误解。而一道选择题只有三个错误选项，最多标出 3 种；实测平均只标出 **2.0 种**。

也就是说：**一道题平均只能让 45% 的已知误解显形，另外 55% 的孩子——无论他脑子里装的是哪一条——在这道题上都不留下任何痕迹。**

把范围收窄到题目较多、统计更稳的 101 个知识点，情况更糟：这些知识点平均记录了 7.7 种误解，单题仍然只标出 2.0 种，**不可识别率 74%**。

**第二个数字：把两个孩子分开，比想象中难得多。**

更严格地问一次：同一个知识点上任取两种误解、任取一道该知识点的题，这道题能不能让这两个孩子给出**不同的**答案？

要能分开，这道题必须同时为这两种误解各准备一个错误选项。实测结果是——**只有 7% 的情况能分开。**

反过来说：**在同一个知识点上，任取两个持不同误解的孩子、任取一道题，超过九成的情况下，这道题看不出他们有什么不同。**

这个 7% 是保守的下界：一个孩子的误解即使没有对应的选项，他也可能碰巧选了别的错项而暴露自己。所以真实的分辨率比 7% 高一些。但它高不到哪儿去——**因为分辨力的上限是硬的：三个错误选项，最多三种误解。**

**第三个数字：大多数误解，整个题库里只有一道题测得出。**

在这 1604 种被记录的误解里，**52.7% 只有一道题能测出来**，74% 不超过两道。

请把这句话在课堂上翻译一遍：如果你这个学期恰好没用到那一道题，那么班上任何一个持这条误解的孩子，从头到尾都会以“正确”的面目通过——**不是他藏得好，是没有任何一道题问过那件事。**

**还有一个顺带的发现。**这 1869 道题里，只有一半（50.2%）的三个错误选项**全部**对应到了已知的误解。另一半题目里，至少有一个错误选项是“没有人知道为什么会有人选它”。

**这些数字改变了什么？**

它把第七章的第一阶段从一个说法变成了一个可以查的事实：**不可识别不是偶尔发生的意外，它是常态。**而且它不是老师的失职——出题的人也没办法，一道四选一的题，天生只有三个位置。

所以那句备课判断要说得更硬一点：

**不要指望一道题能诊断一个知识点。诊断的最小单位不是一道题，是一组互补的题——它们的错误选项加起来，才覆盖得住这个知识点上的那几种误解。**

## 四、四个阶段：正确性消音循环

先把三门学问的词收在一张表里。

| 这个词  | 一句话解释           | 在这一章里对应什么  |
|------|-----------------|------------|
| 可识别性 | 给再多同类数据也分不出两个模型 | 这道题分不出两个孩子 |

|       |                 |              |
|-------|-----------------|--------------|
| 观测设计  | 换一种测法，让两者给出不同结果 | 有区分力的题       |
| 预测误差  | 结果和预期之间的那个差     | 学习真正发生的地方    |
| 预期    | 必须先有它，才可能有落差    | 让孩子先说        |
| 方向性线索 | 能显著缩小答案空间的提示    | 会顶掉预期的那种话    |
| 信息级联  | 跟随前人，自己的线索被弃用   | 全班跟着优等生举手    |
| 私人信号  | 你在看见别人之前的那份判断   | 被挤出公共空间的三十九份 |

三家争的是谁先动。而在真实的课堂里，**没有固定的起点。**

它是一个圈，四步走完回到原处：

**第一步，不可识别。**任务、图形和提示过于同质，正确的规则和错误的规则产出相同的答案。

**第二步，无落差。**老师、同伴、教材或者 AI，在孩子明确形成预测之前给出了方向，结果和旧规则之间的差异无从产生。

**第三步，假一致。**公开的正确被当成多份独立证据，后面的人压下自己的疑问，跟着已有的序列走。

**第四步，再标准化。**老师和系统根据表面上的掌握，缩窄任务、加快进度、把提示提前——**于是下一轮更加不可识别。**

第四步回到第一步，圈闭合了。

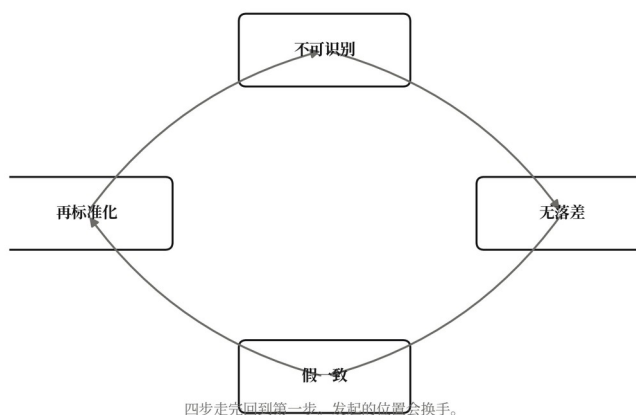


图9 正确性消音循环。

**发起的位置会换手。**在一个高度标准化的题库里，任务先动；在一个 AI 深度介入的个别化系统里，帮助先动——系统一看见停顿就给提示；在一个强调抢答和排名的班级里，公共信号先动；在一次考试复习中，孩子自己的路径可能先动——他一见陌生表示就停，老师为了赶进度给模板，公共正确增加，进入下一轮。

所以诊断不能只问“是不是提示太多”或者“是不是题目变化不够”。**要问的是：这一轮，最先动的是哪一个？**

- 最先改变的是题目的分布吗？

- 最先改变的是孩子的预测吗？
- 最先改变的是公开答案的结构吗？
- 第三个改变之后，下一轮谁先动？

回到那张格子：

|        | 全班都对了                | 有人做不对                |
|--------|----------------------|----------------------|
| 差异还活着  | 活的正确：正确出现在预测、分歧和比较之后 | 正在发生：还没对，但不同的解释正在被区分 |
| 差异已经消音 | 安静的正确：全班一致，没有人再有疑问   | 普通的不会                |

现在，请你回到序言里那三句话。这一章是它们第一次被完整兑现：

**第一，短期指标改善。**课堂更快，正确率更高，孩子更少卡顿，家长更满意。

**第二，失效不发出信号。**错误的规则在当前题型中不暴露，私下的疑问不进入任何记录。你手上的所有数据，都在告诉你一切正常。

**第三，机制自我加固。**越顺畅，越愿意复制同一套任务和提示；越复制，下一轮越顺畅。

这一章说的不是有人做错了什么。每个局部都在解决问题：题目提供了支架，老师及时反馈，孩子参考同伴，课程追求效率。可这些局部的好，叠加起来的结果是：**学习所需要的那个差异，没有被处理，而是被消音了。**

## 五、五个场面，五次差异被消音

（以下场景为构想，内容均为常见教学单元。）

### 场面一：练得越熟，错误的规则可能越牢

分数比较： $\frac{3}{5}$  和  $\frac{3}{7}$  哪个大？

有一类常见的错误规则是“分母大的分数大”。如果这一课的练习恰好都是分子相同、分母不同的题，这条错误规则会**一直给出错误答案**，很快被纠正。

但如果练习是分母相同、分子不同的题（ $\frac{3}{5}$  和  $\frac{4}{5}$ ），那么“分子大的分数大”这条规则——它在这里恰好正确，在别处会出事——就会被大量正确反馈**反复加固**。

**练习不是中性的。它加固的是那个当前有效的规则，不管这条规则将来对不对。**

### 场面二：“会移项”可能只是会跟着符号走

解  $3x + 5 = 20$ ，孩子把 5 移到右边变成 -5，得到  $3x = 15$ 。

正确规则是：等式两边同时减去 5。

错误规则是：过桥变号——搬过去就改符号。

这两条规则在几乎所有的常规方程里给出相同答案。它们的分岔点在哪儿？在  $20 - x = 8$  这样的式子上，在含分式的方程上，在需要说明“为什么可以这样做”的时候。

**如果一整个单元里没有一道题让这两条规则分开，那么这个单元教的就是过桥变号。**

### 场面三：关键词越稳定，关系越难出现

正比例的练习题里，“速度一定”“单价一定”“效率一定”这些短语出现得极其规律。

孩子于是形成一条规则：看见“一定”两个字，就用正比例。

这条规则在整个单元里正确率极高。它在单元测验里的表现，和真正理解了比值恒定的孩子完全一样。

### 场面四：分类讨论被学成“老师提醒后才分类”

含参数的问题里，通常在参数明显出现的地方要求讨论。于是孩子形成一条规则：看见字母就分类。

真正该学的是：什么情况下这个式子会退化，从而必须分情况。这两件事在教材的题里同时成立，所以永远分不开。

而当孩子没发现分岔点时，老师会立刻提醒一句“系数可能为零”——**落差被顶掉了**。接着优等生给出两类讨论，全班跟随——**私人判断退场了**。后面的练习继续把参数放在显眼的位置并标注“分类讨论”——**下一轮更加不可识别**。

一道题，四个阶段全走完了。

### 场面五：标准图形把空间判断变成了视觉模板

三角形的高。教材里的图形几乎总是底边水平、高在内部、垂足看得见。

孩子形成的规则是“从上面那个点向下面那条边画一条竖线”。这条规则在所有标准图里都对。

它的分岔点是钝角三角形——高落在图形外面。而这类图，通常出现在这个单元最后的两道“提高题”里。

**九成的练习加固错误规则，一成的练习负责纠正它。这个比例本身就说明了问题。**

## 六、三个可以真的记录的读数

这一章最容易变成一句抱怨。所以我给三个能记录、不用猜孩子心思的东西。它们看的都是同一件事：**差异活了多久**。

**读数一：模型分辨率。**

在你这一节课的练习里，有几道题能让目标规则和至少一种常见错误规则给出不同的答案？

如果一道都没有，那么这节课的正确率再高，也不含诊断信息。

**读数二：预测先行率。**

在孩子第一次写下或说出他的判断之前，他有没有先拿到任何方向性线索？

这个数很好记：**一节课里，有多少次是孩子先说，多少次是提示先来。**

**读数三：私人判断的保留率。**

同一道题，有多少个孩子的答案是在看见别人的答案之前给出的？

一个几乎零成本的做法：**让全班同时写在纸上，再一起亮出来**，而不是一个一个举手。就这一个改动，能把一节课的独立判断从 1 份变回 40 份。

## 七、这套眼光的分寸

**第一，这一章不反对正确。**

问题不是正确出现了，而是**正确在什么之后出现**。如果孩子先提出了预测，任务让不同的规则分岔，私人判断在公共答案之前留下了记录，然后正确结论在比较中形成——**这个正确是学习的结果**。只有当正确先于这些过程出现，它才成了学习的替代品。

**第二，不是所有的提示都会消除学习。**

情绪支持要给，任务澄清要给，卡住太久要接住。**这一章只针对一种：在孩子形成判断之前给出的方向性线索**。而且只要求你在关键的那一两道题上守住时机，不是每道题都憋着不说。

**第三，不是所有的错误都值得保留。**

抄错数字、写漏符号，就直接指出来。**值得慢下来的，是那种“由一条规则产生的、可以被定位的错误”**。区分这两种，是老师的专业判断。

**第四，专家的快速正确不是消音。**

一个熟练的人一眼看出答案，是因为他的模型已经稳定，不是因为差异被压平了。判断的依据不是快慢，是**换一个结构他还行不行**。

**第五，AI 会让这个循环转得更快，除非你改一件事。**

方向性线索的等待成本几乎归零了——孩子一停顿，提示立刻就到。AI 能生成无限多的同类题，但**同类题不会提高分辨率**。AI 的语言极其流畅，这本身就是一个很强的公共信号，会比班上任何一个优等生更快地建立级联。而数据看板会把这一切显示为“掌握度上升”。

要改的只有一件事，而且很便宜：**让孩子先写下他的预测和理由，再让 AI 说话**。顺序变了，同一个工具的作用就反过来了。

### 这一章的一句话

一个班级全部答对的时候，它可能拥有四十份理解，也可能只拥有一份——而这两种情况在成绩册上长得一模一样。

下一章讲第二个原因。它和这一章无关，却指向同一个结果：**当一间教室被安排得非常高效的时候，学习需要的那点余地，会被最先清掉。**

## 第八章·课堂越高效，越难发生

第三编 为什么它这么少发生·之二

### 一、四十分钟里，没有一分钟是空的

下面这个场面是我构想的，但它就是今天大多数常态课的样子。

一节四十分钟的数学课，教案是这样安排的：

导入 3 分钟，复习旧知 4 分钟，例题讲解 8 分钟，同类练习 6 分钟，变式练习 6 分钟，小组讨论 5 分钟，展示交流 4 分钟，总结 2 分钟，布置作业 2 分钟。

加起来正好 40 分钟。每一分钟都预先有用途。

这是一份好教案。它结构清晰、层次分明、时间精确，在任何一次评课里都不会被挑毛病。

现在，第 22 分钟，一个孩子举手说："老师，我用了另一种方法，但是我算出来的不一样。"

这一刻，教室里发生了什么？

如果他的方法是对的、只是算错了，需要 2 分钟。

如果他的方法揭示了一个班上很多人都有的误解，需要 6 分钟。

如果他的方法其实指向了下一个单元的内容，需要 10 分钟，而且今天讲不完。

而这份教案里，**没有这 2 分钟**。

老师大概率会说："这个想法很好，我们课后再讨论。"——这句话在中国的课堂上一天要被说几万次，而且说的时候完全是善意的。

课后当然不会讨论。因为下一节课也是四十分钟，也每一分钟都排满了。

我要说的不是这份教案糟糕。恰恰相反：**正因为它足够好、足够满、足够高效，那个举手的孩子才没有地方可去。**

### 二、这一章要还的债

数学学习很少发生，是因为课堂被安排得太满了——满到没有一处余地，可以让一个还没结算清楚的差异活到下一轮。

这里说的"余量"不是自由活动时间，也不是"少布置点作业"。它指的是一件很具体的事：一个还没有被处理完的差异，能不能活着走到它可以被处理的那一刻。

### 三、三门学问，三种"余量是怎么没的"

#### 排队论：接近满载的系统，容不下波动

先说清楚排队论在这里的那条结论。

排队论研究的是：任务不规则地到来、每个任务花的时间又不一样，这时候系统什么时候开始堵。

先说一个几乎所有人都体验过、却很少想明白的现象：**高速公路上并没有事故，车速却突然降到二十码。**

原因不是车太多，是车**太满**。车流量在七成以下的时候，一个人点一脚刹车，后面的人稍微松一下油门就吸收了；到了九成五，同样一脚刹车会向后传成一堵墙。

排队论把这件事说得很精确：**当利用率接近百分之百，等待时间不是缓慢上升，是急剧飙升。**而且还有第二条，同样重要：**波动越大，同样的服务水平下，能安全维持的利用率就越低。**

三者不可能同时拿到：**高利用率、大波动、低等待——只能挑两样。**

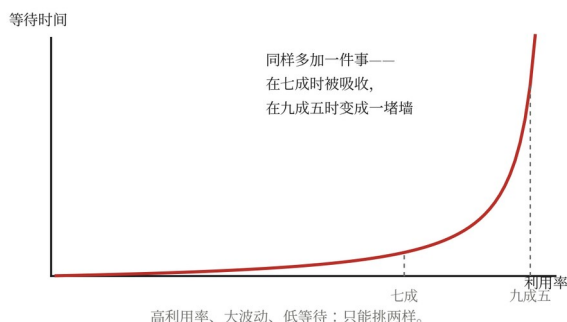


图 10 利用率一旦接近满载，等待时间不是缓慢上升，是急剧飙升。

医院手术室是最典型的例子。管理者希望手术室的排程满一点，因为空着就是浪费。可手术时长天然有波动——同一个手术，快的一小时，慢的三小时。排得越满，后面的病人等得越久，超时越离谱。所以真正专业的排程，**会主动留出空档。**

搬进教室，不需要把孩子比作车或者病人。只需要承认一个事实：**一节课的时长是固定的，老师的注意力和公共讲解的通道是有限的，而不同的孩子完成同一个认知任务所需要的时间不一样。**

只要课堂允许真实的思考，处理时间就一定有波动。

在利用率不高的系统里，两分钟的波动可以被余量吸收。

在接近满载的系统里，两分钟会被读成“课堂失控”。

**而排队论自己最清楚：不是所有的空闲都是浪费。**一个要吸收波动的系统，必须保留与波动程度相匹配的余量。“满载”不是自然事实，它是一项选择。

排队论研究的不是“排队很烦”。它研究一个更硬的问题：当任务以不完全规则的间隔到达、每个任务需要的处理时间又有波动时，系统在什么时候出现等待、拥堵和崩溃。

结论里最重要的一条是：**当利用率接近百分之百，等待时间会急剧增大。**而且波动越大，在同样的服务质量下，系统能安全维持的利用率就越低。

**高利用率、过程波动、低等待——这三件事不可能同时免费拿到。**

把这条搬进教室，不需要把孩子比作流水线上的零件。只需要承认一个事实：**一节课的时长是固定的，老师的注意力和公共讲解的通道是有限的，而不同的孩子完成同一个认知任务所需要的时间不一样。**

有的孩子第一眼就认出了对象，有的要重新画图；有的用算术关系，有的想列方程；有的第一次假设就接近关键结构，有的必须先走一条错路。

只要课堂允许真实的思考，处理时间就一定有波动。

在一个利用率不高的系统里，两分钟的波动可以被余量吸收。

在一个接近满载的系统里，两分钟会被读成“课堂失控”。

于是老师面临一种并非出自观念的压力：要保持公共通道流动，就必须减少个体处理时间的差异。而最直接的办法，不是让每个孩子获得更好的内部组织——那太慢了——**而是让所有人使用同一条外部路径**：同一张图、同一句口诀、同一个设问顺序、同一套板书格式。

题目出现之后，孩子不必决定先看什么，因为老师已经划了线；不必判断用什么单位，因为关键词已经标出；不必比较候选，因为例题类型已经命名；不必决定何时放弃旧路，因为讲解的节奏会准时推进。

**课堂因此越来越像一个低方差系统。它确实高效。**

**但排队论自己知道一件事**：不是所有的空闲都是浪费。恰恰相反，一个服务系统要吸收波动、维持稳定，就必须保留与波动程度相匹配的余量。“**满载**”**不是一个自然事实，它是一项选择。**

而且，为什么要选择满载？排队论自己不回答。容量目标、完成率、时间预算，都来自组织的记录和问责制度。

## 行为生态学：被看见的迟疑，会变成风险

**先说清楚行为生态学在这里研究什么。**

它研究动物怎样在两件相互冲突的事之间分配行动：**去找吃的，和别被吃掉。**

一只小鱼要不要游到那片没去过的水草里？那儿可能有食物，也可能有捕食者。它的决定不只取决于“那儿有没有好东西”，还取决于**这一带的危险程度**。

有一类实验做得很干净。特立尼达岛的溪流里，同一种孔雀鱼生活在两类环境里：一类河段捕食者多，一类几乎没有。把两地的鱼捞来做同样的探索测试，结果稳定地不同——**高风险环境里长大的鱼，面对不确定的新线索时，探索的范围明显更保守**；而且高风险群体里，很快就会分化出“带头的”和“跟着的”，个体之间的行动差异反而变小，整群挤得更紧。

请注意这句话的结构：**它们不是不好奇，是它们非常准确地读懂了环境。**

行为生态学由此给出一条可以检验的结构：

**当探索同时带来信息和暴露成本时，行动者不只根据“能不能发现新东西”决定要不要探索，还要根据“偏离安全区会不会被看见、被落下”来调整搜索范围。**

搬进教室，两个词要换一下：

- **庇护**是什么？是老师刚示范的方法，是同桌已经确认的答案，是题目里的关键词，是全班同步前进的节奏。
- **暴露**是什么？不是答错。**是让别人看见你还没跟上共同的速度**——你问了一个大家好像都懂的问题，你画了一张没人用的图，你的路径需要全班多等一分钟。

**这个理论也自带警告**：风险不会机械地减少所有探索。线索是否可靠、群体结构如何、资源紧不紧张，都会改变反应；有些高风险环境里的种群反而演化出更高的活动性。**所以“风险”不是一个固定的刺激，它取决于系统怎样组织信息和庇护。**

排队论解释了老师为什么难以容忍差异，但它没有解释另一个更常见的现象：**老师明明没有催，孩子也会主动放弃探索。**

一道开放题刚出现，孩子先看老师的表情；同桌已经写下答案，他就擦掉自己的图；老师问“还有不同方法吗”，全班短暂沉默，然后有人重复了板书上的路径。

没有命令，没有惩罚，探索却迅速收缩了。

行为生态学研究动物怎样在获取资源和避免危险之间分配行动。探索能带来新信息、新食物、新空间，但也意味着离开熟悉的庇护、暴露在捕食者面前。**探索从来不是无条件的美德，它是一种被风险结构调节的行为。**

有实验发现，在高捕食风险环境里的鱼，面对模糊的新异线索时，探索空间的均匀程度会下降；也有研究发现，高风险环境里的鱼群会更快地分化出发起者和跟随者，个体行为的差异反而变小，群体更紧密。

这些当然不能直接证明孩子把教室当成了捕食现场。但它们给了一条可以检验的结构：

**当探索同时带来信息和暴露成本时，行动者不只根据“能不能发现新东西”决定要不要探索，还要根据“偏离安全区会不会被看见、被落下、被否定”来调整搜索范围。**

数学课堂里的**庇护**是什么？是老师刚示范的方法，是同桌已经确认的答案，是题目里的关键词，是“这类题都这么做”的标签，是全班同步前进的节奏。

数学课堂里的**暴露**是什么？不是简单的答错。**是让别人看见你还没有跟上共同的速度：**你问了一个大家好像都懂的问题，你画了一张没人用的图，你在老师准备收束时还坚持另一个假设，你的路径需要全班多等一分钟。

**满载的课堂把“多想一会儿”变成了一个公共事件。**一旦迟疑既可见又有代价，它就会被孩子读成风险信号。

于是孩子做出一种非常理性的调整：少走远路，少提出还不成熟的猜想，优先复制说话最有分量的那个人的策略，在模糊的地方等公共信号，而不是自己去找证据。

**这不是“孩子没有好奇心”。**恰恰可能因为他极其敏锐地读懂了环境：**这间教室奖励的不是探索带来的信息，而是尽快回到安全路线。**

## 审计学：没有进入记录的，在组织里等于没有发生

先说清楚审计学在这里的那个洞察。

审计的存在理由很朴素：一个组织不能只相信当事人说“我干得不错”，必须留下能被**第三方检查的证据**。

这件事本身没有问题。问题出在第二步——**什么东西能进入证据系统，会反过来改变人们做什么。**

这个现象有两句著名的话：

- **古德哈特定律：**一旦一个指标成为目标，它就不再是一个好指标。
- **坎贝尔定律：**一个社会指标被用于决策的分量越重，它越容易被操纵，也越容易扭曲它本来要衡量的东西。

审计研究里有一些很具体的实验支持这件事。比如：给审计人员一组被“**基准化**”的绩效指标，他们的注意力会集中到这些指标上，而**忽略那些没有被基准化、却在各个部门普遍存在的问题**。还有一个更微妙的：把内容完全相同的工作底稿交给复核者，只改一个信息——这份底稿完成得比预算时间**快还是慢**——复核者对质量的评价就不一样了。

一句话记住：**记录不是活动之后的中性镜子。记录选择什么，活动就会向什么靠拢。**

数学课堂最容易记录什么？完成了多少页、正确了多少题、多少人达标、作业是否按时提交。

最难记录什么？**一个孩子今天放弃了哪条错误规则；他怎样在两种单位之间重新选择；他的沉默到底是空白还是在比较两个模型。**

后者不是绝对不能记录，但成本高、编码有争议、第三方难以复核。前者整齐、便宜、容易比较。

于是资源会流向可审计的产出。**不是因为老师虚伪，也不是因为管理者愚蠢，而是因为组织层面，“无法被稳定复核的好”，很难跟“根本没有发生”区分开。**

老师可能愿意等，孩子也可能愿意探索。但学校仍然要回答一串现实问题：这节课完成了什么？学生掌握到哪里？进度是否达标？谁需要补救？

这些问题不能靠感觉，需要证据。

审计学研究组织怎样通过文件、抽样、指标和复核，把复杂的活动变成可以追踪、可以问责的记录。**审计并不天然敌视真实质量。**它的必要性来自一个朴素的事实：组织不能只相信一个人说“我做得很好”，必须留下能被第三方检查的证据。

**困难在于：什么能够进入证据系统，会反过来改变人们做什么。**

有实验发现，外部基准会让审计人员集中注意那些被基准化的指标，而忽略没有被基准化、却普遍存在的指标；还有实验发现，即使工作底稿的内容完全相同，只要完成时间低于或者高于预算，复核者对质量的判断就会不同。

**一句话：记录不是活动之后的中性镜子。记录选择什么，活动就会向什么靠拢。**

数学课堂最容易记录什么？完成了多少页，正确了多少题，多少人在规定时间举手，多少人达标，作业是否按时提交，知识点是否覆盖。

**最难记录什么？一个孩子今天放弃了哪条错误规则；他怎样在两种单位之间重新选择；他的沉默到底是空白还是在比较两个模型；一次没做完的推导为下周留下了什么；一场没有结论的争论改变了谁的提问方式。**

后者不是绝对不能记录，但成本高、编码有争议、第三方难以复核。前者整齐、便宜、容易比较。

于是资源会逐渐流向可审计的产出。**不是因为老师虚伪，也不是因为管理者愚蠢，而是因为组织层面，“无法被稳定复核的好”，很难跟“根本没有发生”区分开。**

要证明工作完成了，就得留下完成的痕迹；要让痕迹可比较，任务就得标准；要按时生成标准痕迹，路径就得缩短，停顿就得减少。

**而审计学自己也最清楚：**指标会产生博弈。被测量的人不只是服从指标，还会调整报告、选择样本、重排时间，让指标看起来改善。所以可审计性也不是外面强加的静态制度——它会被它所测量的行为反向塑造。

## 四、四个阶段：余量清零循环

先把三门学问的词收在一张表里。

| 这个词     | 一句话解释            | 在这一章里对应什么      |
|---------|------------------|----------------|
| 利用率     | 系统被占用的比例         | 课表排得多满         |
| 波动      | 每个任务花的时间不一样      | 不同孩子需要的时间不同    |
| 余量      | 用来吸收波动的空档        | 让一个差异活到下一轮的空间  |
| 探索—安全权衡 | 找新东西和别被吃掉之间的取舍   | 要不要说出那个还没想清的想法 |
| 庇护／暴露   | 待在安全区／被看见偏离了共同速度 | 跟着标准步骤／多要一分钟   |
| 可审计     | 能被第三方按稳定规则复核     | 能写进成绩册的东西      |

|        |                 |                |
|--------|-----------------|----------------|
| 古德哈特定律 | 指标一旦成为目标就不再是好指标 | 正确率越好看，越说明不了问题 |
| 指标反应性  | 被测量的人会调整行为去迎合测量 | 课堂向“能留下痕迹的路”靠拢 |

三家都在争谁先动。排队论说只要利用率降下来，别的都好办；行为生态学说只要探索不再危险，孩子自然会用余量；审计学说只要承认未完成的过程也算证据，时间和风险都会重新分配。

三家不能同时是第一因。真实的情况是——它是一个圈，而且可以从三个不同的门进去。

**第一步，满载。**课程、注意力和公共通道被安排到接近没有空档，个体处理时间的差异开始形成积压。

**第二步，避险。**积压让迟疑和绕路变得可见，孩子把偏离共同速度读成风险，转向安全路径。

**第三步，可审计。**安全路径生成整齐、按时、可比较的表现；那些无法即时结算的过程，退出了正式记录。

**第四步，再满载。**整齐的记录被解释为“课程容量还够、标准路径有效”，于是下一轮进度更紧、提示更早、任务更多。

第四步回到第一步。

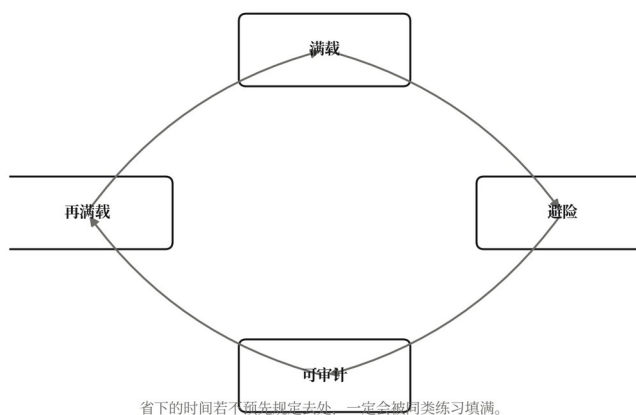


图 11 余量清零循环。

**发起的位置照样会换手。**有的班先被进度挤满——老师怕讲不完，提前给提示，孩子学会等提示，记录系统最后把顺畅完成认定为有效教学。有的班先形成公开风险——一次嘲笑、一次排名、长期的速度比较，让孩子不再公开探索；路径趋同之后课堂更容易按时完成，管理者随后压紧进度。有的学校先改变评价——平台开始追踪完成率，老师为了生成数据而缩短过程，孩子随后学会只走能留下痕迹的路。

回到那张格子：

|  |         |        |
|--|---------|--------|
|  | 这节课完成度高 | 这节课没上完 |
|--|---------|--------|

|          |                       |                       |
|----------|-----------------------|-----------------------|
| 差异活到了下一轮 | 活的正确：有东西没结算完，而且它进了下一轮 | 正在发生：进度乱了，但有一个真问题被留住了 |
| 差异当堂清零   | 高效空心：快、准、齐，什么也没长出来    | 普通的失控                 |

第七章那三个性子，在这里一条不落地重演了一遍：**课堂更快更齐，评课分更高；没有任何一个数据字段叫“今天被清掉的差异”；而上一轮正确率提高之后，下一轮不是撤支架，是加任务量。**

## 五、四个场面，四次余量被清掉

（以下场景为构想，内容均为常见教学单元。）

### 场面一：竖式越整齐，位值关系越容易退场

两位数乘法。竖式是一个极其高效的算法：格式固定，步骤明确，容易批改，错误容易定位。

代价是：那个“用十位上的2去乘，得到的其实是20个”的位值关系，被压进了“第二行要往左错一位”这个格式约定里。

**孩子照着做就一定对。而正因为一定对，那个关系没有机会出场。**

省下来的时间去哪儿了？通常是再做十道竖式。

第四章那个数字在这里可以再读一次：一千多节课的录像编码显示，本意在于“建立联系”的题目，最后真的建立起联系的，在美国不到1%。**那些题不是被删掉的，是在一节按时完成的课里被磨平的。**

### 场面二：“讨论三分钟”是怎么变成表演的

小组讨论被写进教案，限定三分钟。

三分钟能干什么？能让每个人说一句话，能让组长汇总，能形成一个可以展示的结论。

**三分钟不足以让两个真正不同的想法撞上，因为撞上之后需要的时间是不可预测的。**

于是孩子很快学会了三分钟的正确用法：先看谁的想法最容易被接受，然后一起把它整理得漂亮一点。**讨论变成了整理，因为只有整理的时间是可控的。**

### 场面三：视觉模板用低等待换走了空间判断

三角形的高，第七章我们从“不可识别”的角度看过。这一章看的是另一面：**为什么老师会给出那个标准图形？**

因为不标准的图形会让一部分孩子卡住，而卡住的时间不可预测。给一张标准图，全班的处理时间立刻收敛。

这是一个理性的选择。它的代价是那个“高是一段垂直距离，不是图上的一条线”的判断，被一张图替掉了。

### 场面四：图像生成器可能把思考的余量压到零

函数教学里，画图曾经是最花时间的一步——描点、连线、画错、重来。

现在有软件，输入解析式，图像立刻出来。省下的十分钟可以多讲两个题型。

**但那十分钟里原本会发生的事情不见了：**为什么这里陡、那里平；为什么点越取越密的时候形状开始稳定；为什么这个点算出来和图上看到的对不上——**这些疑问只在“慢慢画”的过程里出现。**

工具本身没有错。**错的是我们默认省下来的时间应该拿去多讲两个题型。**

## 六、三个可以真的记录的读数

**读数一：余量存活时长。**

一个还没有被处理的差异，从它出现到它被处理或者被放弃，中间隔了多久？

在大多数课堂上，这个数是**零到三十秒**。

**读数二：差异回写率。**

这节课上出现的分歧、疑问和错误里，有多少**具体地改变了下一节课的某一个条件**（哪一道题、哪一句提问、哪一个例子）？

不是“我记住了”，是“下一轮真的变了”。

**读数三：不可审计证据的转化率。**

那些没法写进成绩册的东西——一次没做完的推导、一次被否掉的猜想、一次沉默的比较——**有多少最后进入了某种正式的记录**（哪怕只是你教案边上一行字）？

如果长期是零，那么在这套系统里，它们等于从未发生过。

## 七、这套眼光的分寸

**第一，效率不是敌人。**

一节松松垮垮、四十分钟只讲了两道题的课，不会因为“有余量”就更好。**这一章说的是余量的用途，不是余量的多少。**一节高效的课如果留出了三分钟给一个真问题，它比一节拖沓的课强得多。

**第二，余量也会被浪费。**

给了时间，不等于差异会被处理。有的课堂留出了空白，孩子拿去发呆或者聊天。**余量必须有去处**——它得通向一个具体的东西：一个要回答的分歧、一个要设计的反例、一条要写进下一轮的规则。

**第三，可审计不是坏事。**

第三方能够复核，是公平的基础。没有记录，一个孩子的进步就只能由老师个人说了算，这对孩子并不安全。**要改的不是“要不要记录”，而是“能不能给那些难记录的东西也留一个字段”。**

**第四，这不是在指责哪个老师。**

一个老师在满载的课表里选择压平差异，是在他的约束下的正确选择。**这一章说的东西，多数不在一节课的权限之内。**一个老师能做的，是在他有权限的地方留一处口子；能改结构的人，应该去改结构。

**第五，AI会让清零更彻底，除非你先定好省下的时间归谁。**

AI会自动填满省下的时间；它把迟疑变成可以立刻消除的异常；它扩大了可审计的范围——现在连过程数据都能记录了，**而被记录的过程仍然只是那些能被记录的过程。**

所以在引入任何提效工具之前，先回答一个问题：**这次省下来的时间，归谁？**如果答案是“多做几道题”，那么这次提效的净效果，就是把余量清得更干净。

## 这一章的一句话

| 一间教室最稀缺的东西不是时间，是一个还没算清楚的问题能够活到明天的权利。

下一章讲第三个原因，也是最不容易看见的一个：当数学被切成一个个便于搬运的箱子，那些跨箱的关系，就再也没有地方存放了。

## 第九章·越容易搬运，越难长出来

第三编 为什么它这么少发生·之三

### 一、“这是哪一类题？”

下面这个场面是我构想的，但那句话你一定听过无数遍。

一个孩子拿着一道题来问你。他念完题，第一句话不是“这道题里两个量之间是什么关系”，而是：

“老师，这是哪一类题？”

他不是偷懒。他是在做一件在这个系统里完全正确的事——先确定这道题属于哪个箱子，然后调用那个箱子里的方法。

因为这套系统就是这么运转的。教材按章节分好，练习按题型编好，作业按知识点归好，平台按标签推好。他从上学第一天起接收到的信息就是：**数学是由一个个类别组成的，学数学就是把题目放进正确的类别。**

现在我们看这套系统怎么把一个孩子的问题吃掉。

分数加法。他把  $1/2 + 1/3$  算成了  $2/5$ 。

系统的反应链是这样的：这个错误被归类为“异分母分数加法错误”；平台推送更多通分练习；孩子学会了先通分再加；即时表现改善，这个知识点被标记为“已掌握”。

**但他仍然把分数看成上下两个整数。**

这个问题到了分数乘法会再出现一次，到了比会再出现一次，到了概率还会再出现一次——**而每一次，它都会被归进一个新的箱子：**“分数乘法薄弱”“比的意义未掌握”“概率概念不清”。

没有一个箱子叫“他还没有建立分数单位”。因为这不是一个知识点，**这是一条横穿好几个箱子的关系。**

### 二、这一章要还的债

**数学学习很少发生，是因为便于搬运的那个单位，取得了定义学习单位的权力。**

一个东西要在千万人之间交接——跨教师、跨学校、跨学期、跨平台——它就必须有稳定的边界、清楚的名字、可查的进度。这些要求塑造了“知识点”这个东西。

**而知识点是一个运输单位，不是一个学习单位。**我们长期把这两个混为一谈。

### 三、三门学问，三种“谁来切分知识”

**系统分类学：边界一旦确定，世界才成为可讨论的对象**

**先说清楚系统分类学在干什么。**

它不是“给生物贴名字”这么简单。它要决定：**哪些差异足以把两个个体分成不同的类，哪些相似只是表面相似。**

有一件事外行常常不知道：**生物的分类一直在改，而且改得很厉害。**

过去分类主要看形态——长得像的归一起。分子证据（DNA 测序）进来之后，很多老类群被拆散或者重组：有些看起来很像的物种，亲缘关系其实很远（因为在相似的环境里各自演化出了相似的样子）；有些长得完全不同的，反倒是近亲。

于是这门学科里有一股持续的张力，而且它的行业规则把这股张力明写出来了：

- 一方面，**名称必须稳定**。如果今天的“某种甲虫”和明天的“某种甲虫”指的不是一个东西，所有的观察记录、实验和文献就全接不上了。国际的命名法规追求名称的唯一和稳定，就是为了这个。
- 另一方面，同一份法规又明确规定：**命名规则不应该限制分类思想本身**——当科学理由要求重新分类时，名称和归属可以被修订。

请把这两句话放在一起看。它们合起来说的是：

**分类不是自然界贴在自己身上的标签，是我们为了能一起讨论而钉上去的；它必须稳定到足以承载沟通，又必须可修改到足以服从新的证据。**

这就是这一章最要紧的对照。一门以分类为生的科学，尚且明写着“证据可以改写边界”；而数学课程里的分类——章、节、题型、知识点——却几乎从来不接受回写。

系统分类学处理的不只是给生物贴名字。它要决定哪些差异足以把两个个体分成不同的类群，哪些相似只是表面相似，名称怎样保持唯一和稳定。

国际的命名法规明确追求名称的稳定和普遍，要求每一个分类单元有清晰独特的名字；**同时又明确规定，命名规则不应限制分类思想本身——当科学理由要求重新分类时，名称和归属可以被修订。**

这两句话之间有一股持续的张力。

一方面，没有稳定的分类，知识无法积累：今天的“某种甲虫”和明天的“某种甲虫”如果指向不同的东西，观察和文献就接不上。另一方面，**分类不是自然界贴在自己身上的标签**——新的形态证据和遗传证据会迫使人们重画边界。

如果让分类学来解释数学学习为什么稀少，它会说：**因为课程在关系证据出现之前，就把分类结果交给了孩子。**

数学教材总是按已经稳定的对象组织：数、式、方程、函数、图形、统计、概率。分类本来是学术共同体对长期关系证据的结算结果，**在课堂里却变成了学习的起点。**

孩子第一次接触分数，教材已经告诉他这是“分数”；第一次观察两个量共同变化，标题已经告诉他这是“函数”；第一次处理两个未知量的和与差，题库已经告诉他这属于“和差问题”。

**类别不是事后的整理，它提前分配了注意力。**一旦一道题进了“相似三角形”，孩子就开始找对应边和比例——他不需要再判断这道题该不该找。

**但分类学自己最清楚一件事：**名称稳定不等于关系已经终结。命名法规追求沟通稳定，却明确保留了重新分类的自由。

**如果科学分类本身都允许证据改写边界，那么数学课堂就没有理由把教材的分类当成不可回写的自然事实。**真正的问题不是有没有类别，而是——谁拥有改类的权力，什么证据足以让一道题换箱。

**物流学：能不能交接，先决定东西必须长成什么样**

**先说清楚集装箱到底改变了什么。**

标准集装箱不是“一个大铁盒子”这么简单。它是二十世纪最被低估的发明之一，而它改变世界的方式，恰恰是这一章要讲的机制。

在集装箱之前，货物是散装的。一船货到港，几百个码头工人一件一件往下搬——袋装的咖啡、木箱的机器、成捆的布料，每一件形状不同、重量不同、堆法不同。装卸一艘船要好几天，成本极高，破损和偷盗也多。

集装箱做的事情，是把这一切装进一个统一的外壳。箱子的尺寸、承重、四个角的接口、编号方式，全部标准化（ISO 有一整套标准规定这些）。于是：

**港口、吊机、卡车、火车、仓库，都不必再理解每一件货物的内部结构，只需要理解箱子的接口。**

装卸时间从几天降到几小时，运费降了一个量级，全球供应链才成为可能。

**但这里有一件事，比“效率提高了”重要得多。**

集装箱不只是适应了原有的港口——它逼着港口重建。老码头的水深、吊机、堆场、进出道路全都不适配，于是全世界的港口按箱子的尺寸重新设计；船按箱子的尺寸重新造；卡车和火车的车板按箱子的尺寸定；甚至一些老港口因为改不动，直接衰落了，货运中心转移到新港。

一句话记住：**箱子会重造道路。包装不是被动地承载内容，它反过来决定了以后什么东西容易被运。**

装不进箱子的货物——超长、超宽、形状怪的——从此运费高昂、路线稀少。它们没有消失，但它们变贵了、变难了、变少了。

标准集装箱的革命，不在于让运输更快，而在于它把极其异质的货物包进了统一的外形。

箱内可以是机器、衣服、食品或零件；箱外只有标准的尺寸、承重、接口和编号。港口、卡车、铁路、吊机和仓库不必理解每一件货物的内部结构，只需要理解箱子的接口。

这套机制的关键优点是：**运输系统可以忽略内部差异。**

如果让物流学解释数学学习为什么稀少，它会给出一个和分类学不同的答案：**不是分类决定了教学，是大规模交接的要求，迫使知识被做成了标准包装。**

一本教材要跨地区使用，一套课程要跨教师执行，一项评价要跨班级比较，一个平台要向百万学生分发内容——知识就必须有稳定的接口。

它需要被压成“本节目标”“本课重点”“标准步骤”“常见错误”“五分钟微课”“三道基础题”“一项掌握度”。老师换了，包装能接手；学生转学，进度能迁移；平台换了模型，技能图谱还能继续调用。

**这种包装未必来自错误的教育观，它来自协作的规模。**一个只面对一个孩子的老师，可以跟着这个孩子临时生成的关系改写课程；一个面对百万学习者的平台，必须先把内容做成可调用的模块。

一旦进了箱子，知识的来路就被暂时封存了。孩子看到“通分”，看不到为什么分数单位必须重组；看到“移项”，看不到等式两边同时做了什么变换。**运输接口只需要知道“这件货已到达”，不需要重新经历货物被制造的过程。**

而物流学自己也知道：标准箱不是中性的外壳。标准集装箱不仅适应了公路、铁路和海运，它还促使港口、吊机、船舶、仓库和整个贸易网络围绕自己的尺寸重新设计。

**箱子会重造道路。**

同样，教材为了适应课时而切分内容，平台为了适应数据字段而定义技能，评价为了便于批改而规定解题格式——**这些包装反过来决定了以后哪些数学关系容易被发现，哪些只能留在箱缝里。**

## 行政学：无法被看见的学习，无法进入组织现实

先说清楚行政学在这里的那个概念。

它叫**可识读性**（也译作“可辨识性”）。人类学家詹姆斯·斯科特用它解释了一件事：一个**国家要管理它的人民，必须先把他们变成“看得懂”的样子**。

这听起来抽象，但他举的例子都极其具体：

- **姓氏**。在很多地方，固定的世袭姓氏并不是自古就有的，是被国家推行的——因为没有稳定的姓名，就没法征税、征兵、登记。
- **地籍图**。农民对自己那块地的了解是极其丰富的：哪一角积水、哪一片朝阳、哪块地祖上和邻居有约定共用。可这些没法登记。**能登记的只有：编号、面积、边界、所有人**。于是复杂的实际关系被压成一张能收税的表。
- **林业**。十八世纪德国推行“科学林业”，把杂乱的天然林改造成整齐的、单一树种、同龄、成行排列的人工林——好数、好算、好收。**第一代长得极好。第二代开始出问题：土壤退化、病虫害爆发。德语里后来专门有个词，直译就是“森林死亡”。**

斯科特的结论不是“管理是坏的”。**任何组织只要分配资源，就必须建立名称、类别、表格和指标**。他的结论是更麻烦的一句：

**简化让治理成为可能，也可能抹掉那些支撑实际运行的地方性知识。**

而最要紧的是最后一步：**被简化过的那个版本，会逐渐取代真实的版本成为现实**。地籍图画出来之后，纠纷按图判；林子按行种下之后，林业就按行来教。

搬进课堂：孩子重新划分了对象，却没有对应的指标；他发现两个章节其实共享同一种结构，却无法在进度表上增加一格。**组织看不见这些，不是因为它们没有价值，而是因为它们没有行政地址。**

一个老师可以看见一个孩子的犹豫、草稿和突然转向。一个组织不行——它面对的是大量人员、有限资源和责任链条。

它需要知道哪个年级学到哪里，哪个班完成了什么，哪些学生需要支持，哪些投入产生了效果。要完成这些任务，复杂的实践必须变得**可识别**。

这不是管理者恶意简化。任何组织只要分配课时、经费和师资，就必须建立名称、类别、表格和指标。**问题在于，现实一旦进入行政视野，往往只能以被简化过的形式存在。**

关于国家“可识读性”的研究揭示过这一点：中央规划需要把地方上复杂、连续、依赖情境的实践，转化成规则清楚、可记录、可比较的单位。**简化让治理成为可能，也可能抹掉那些支撑实际运行的地方性知识。**

在数学教育里，这个过程非常清晰：孩子重新划分了对象，却没有对应的指标；他改变了单位，却没有完成预设的知识点；他发现两个章节其实共享同一种结构，却无法在进度表上增加一格。

**组织看不见这些变化，不是因为它们没有价值，而是因为它们没有行政地址。**

而“完成第3.2节”“掌握异分母分数加法”“正确率达到85%”，都有清晰的地址。它们可以被汇总、比较、预警和问责。

久而久之，**能进入记录的东西，就不只是学习的证据，而会反过来成为学习本身的定义。**

**但行政系统也不能凭空发明可识别性。**它依赖既有的名称、标准和表格：没有分类，数据无法聚合；没有标准包装，记录无法跨机构比较。**看起来是第一因的行政需求，材料往往是前两轮已经造出来的类别和货件。**

## 四、四个阶段：知识集装箱循环

先把三门学问的词收在一张表里。

| 这个词    | 一句话解释            | 在这一章里对应什么      |
|--------|------------------|----------------|
| 命名的稳定性 | 名字得稳，否则文献接不上     | 章节和题型必须相对固定    |
| 分类的可修订 | 有证据就该重画边界        | 谁有权把一道题换箱      |
| 标准接口   | 不必理解内部，只需理解外壳    | 知识点、掌握度、进度条    |
| 箱子再造道路 | 包装反过来决定什么容易被运    | 课时和字段反过来决定教什么  |
| 装不进箱的货 | 没消失，但变贵、变难、变少    | 跨箱的关系          |
| 可识读性   | 治理必须先把现实变成看得懂的样子 | 学习必须先变成能记录的样子  |
| 地方性知识  | 支撑实际运行、却登记不了的那些  | 老师看得见、系统看不见的东西 |

三家争的是谁先切。分类学说先有边界才有路径；物流学说许多知识点的边界根本来自数学内部，而来自课时、页码、考试结构和平台接口；行政学说哪些差异被编进课标、哪些模块拿到课时，取决于组织需要追踪什么。

三家谁也当不成永久的第一因。而它们共享着一个更隐蔽的想法：

**数学知识已经是某种可以被切分、装箱和盘点的东西，争议只在于切在哪里。**

这个前提看起来无需讨论——教材当然要有单元，平台当然要有技能。但数学学习最关键的事实恰恰是：**孩子在学习之前，并不一定拥有那个等待被搬运的对象。**

"分数单位"不是一件已经存在于头脑中的货物，它要在分割、比较、重组和表达中生成。"函数"不是一个可以直接交付的概念包。**没有被生成的东西，无法被搬运；而搬运系统只认那些已经成形、有标准接口的东西。**

于是这个循环闭合了：

**第一步，分类。**课程先划定对象、知识点和题型的边界。

**第二步，装箱。**这些边界被做成单元、课时、步骤、练习包和平台标签。

**第三步，可治理。**标准货件生成完成率、掌握度、进度和学生画像。

**第四步，再分类。**这些数据被用来细化知识地图、调整资源、重新标记学生——**下一轮从更紧的边界开始。**

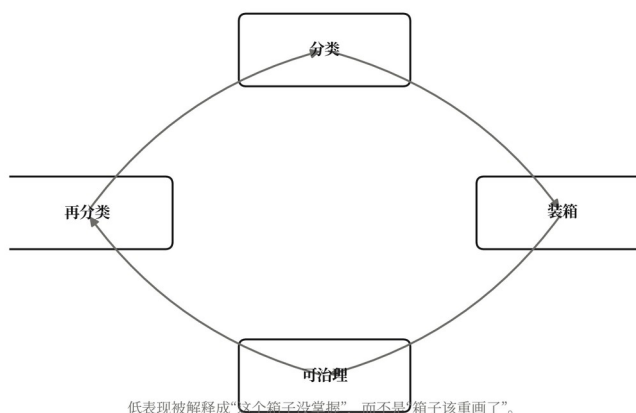


图 12 知识集装箱循环。

发起的位置同样会换手：可以从教材修订开始，可以从一个新的平台数据字段开始，也可以从一次新增的达标指标开始。

回到那张格子：

|        | 在这个箱子里全对              | 在这个箱子里做不对                    |
|--------|-----------------------|------------------------------|
| 箱子能被打开 | 活的正确：他能说出这道题为什么可以归到别处 | 正在发生：还不熟，但他在问“这跟上一章那个是不是一回事” |
| 箱子打不开  | 封箱的正确：箱内全对，出了箱就不认识    | 普通的不会                        |

序言里那三句话，最后一次出现：**短期指标改善**（进度更清晰，正确率更高，教师交接更顺，平台推荐更精准）；**失效不发出结构信号**（孩子在箱外失败，被解释成“原来那个箱子没掌握”，而不是“箱子的边界该重画了”）；**自我加固**（数据越丰富，知识图谱越细；图谱越细，下一轮越依赖同样的标签）。

## 五、五个场面，五个箱子

（以下场景为构想，内容均为常见教学单元。）

### 场面一：“同分母/异分母”是仓库分类，不是关系分类

这两个词把分数加法切成了两个箱子。孩子学会：先看分母，一样就直接加，不一样就先通分。

**但从关系上看，这两件事是同一件事：**把两个分数放进同一个单位再相加。同分母只是“已经在同一个单位里”的特例。

按仓库分类，它们是两类；按关系分类，它们是一类的两种情形。**孩子学到的是仓库的分类。**

## 场面二：“移项”把一次变换做成了运输标签

移项是一个极其好用的标签：可以被检查，可以定位错误，可以统计正确率。

代价是它把“等式两边同时做同一种可逆操作”这个关系，压成了一个动作口令。孩子一旦在复杂的方程里出错，系统的反应是**分配更多“移项变号”的练习**——继续在同一个箱子里加货，而不去问这个箱子的边界对不对。

## 场面三：“鸡兔同笼”把一种结构变成了地方特产

假设法是一种非常一般的推理结构：先把对象统一，再用差异恢复对象的差别。

它可以用在鸡兔同笼，也可以用在盈亏问题、浓度问题、平均数问题、方程组的消元。

因为它进了“鸡兔同笼”这个箱子，孩子会觉得这是一道有名字的怪题，需要专门的技巧。（第一章我们从另一个角度看过这道题：那时候关心的是“先假设全是鸡”这条被排除的路有没有留下来；这里关心的是这条推理结构为什么被一个题目名字锁住了。）**箱子的名字太响亮，把结构的普遍性盖住了。**

## 场面四：函数被章节切成了三种互不相干的产品

正比例在一处，一次函数在另一处，反比例在第三处，中间隔着别的内容。

对课程管理来说，这样切很合理：每一块都有清晰的目标、课时和测验。

对孩子来说，“共同变化”这一件事，被切成了三种要分别记住的东西。**他很难自己发现它们是同一个问题的三种情形——因为课程从来没有给他提供把三个箱子并排看的场合。**

## 场面五：模型库越完整，关系的缺口越容易被盖住

几何辅助线的“模型库”是近年最流行的箱子：一线三等角、手拉手、半角模型、将军饮马……

它们确实有用。但一个只有模型库的孩子，遇到新图形时会去检索“这像哪个模型”，而不是去问那五个真正的问题：**原图里哪一种关系不可见？证明目标需要哪一种新关系？连接哪些对象会产生这种关系？新增的复杂性值不值？哪个近邻图形会让这条线失效？**

**模型库越完整，他越不需要问这五个问题。**

## 六、三个可以真的记录的读数

**读数一：拆箱重构率。**

给一道**不带任何标签**的题——没有章节位置，没有题型名，没有平台标签——孩子有多大比例能自己判断出它跟什么有关？

最省事的做法：期末把不同章节的题混编，去掉标题。**这一份卷子和按章节考的卷子之间的分差，就是箱子替他做了多少事。**

**读数二：分类回写率。**

孩子提出的新分组——“这道题跟上一章那道其实是一回事”——有多少真的改变了后续的任务、标签或者顺序？

如果长期是零，那么这个课堂里，改类的权力完全不在学习者手上。

**读数三：双向边界辨识度。**

不只问“这道题属于哪一类”，还要问：“**这道题为什么不属于那一类？**”

第二问几乎从来没有人问。而边界是双向的——只知道进去的路，不知道出去的路，那不是边界，那是入口。

## 七、这套眼光的分寸

**第一，章节边界经常是对的。**

数学内部确实有真实的结构差异。整数和分数、代数和几何的区分不是人为发明的。这一章反对的不是分类，是“分类只能单向施加、不能被证据回写”。

**第二，标准化让公平成为可能。**

统一的教材和评价，让一个偏远地区的孩子和大城市的孩子面对同样的内容。取消标准化不会让学习变好，只会让不平等变大。要争取的是在标准之内保留一条改类的通道，不是拆掉标准。

**第三，大规模系统不可能实时接受所有的改类请求。**

一个平台不能因为一个孩子说“这两道题是一回事”就重排知识图谱。现实的做法是留一个入口，定期处理——比如一个学期整理一次孩子们提出的“跨箱关系”，看看有没有一条值得写进下一轮的教研材料。

**第四，孩子提出的新分组不天然正确。**

“这两道题都有3”不是一条关系。改类的请求必须承担举证责任（第五章那套东西在这里正好用得上）：你说它们是一回事，请指出共享的是哪一条关系，以及什么反例会推翻它。

**第五，AI 可能把箱子做得更小更牢，也可能成为拆箱的工具。**

它天然倾向于前者：无限细分知识点，自动标注，把暂时的分类固化成基础设施，用流畅的对话把运输接口藏起来。

但它也是我们第一次拥有的、真正廉价的拆箱工具。只要改一个用法：不要问它“这是哪一类题”，而是问它——

“这道题还能被归到哪些别的地方去？分别是因为什么关系？”

这个问法，AI 答得比任何一本教辅都好。而它恰好是这一章想要的那个动作。

## 这一章的一句话

一个孩子问“这是哪一类题”的时候，他其实是在问一个仓库问题。而数学从来不是一个仓库——它是一张关系的网，箱子只是我们为了把它搬来搬去而临时钉的。

第三编到这里结束。三章给了三个答案：正确来得太早（第七章）、余量被清零（第八章）、运输单位取代了学习单位（第九章）。

它们看起来毫不相干：一个说的是任务和提示，一个说的是时间和风险，一个说的是分类和标签。

但你应该已经认出来了——它们是同一张脸。

三个都是自我加固的圈；三个都在短期指标上表现为改善；三个失效的时候都不报警。而且三个里面，没有任何一个是因为有人做错了什么。

这就是接下来的结语要处理的问题。

## 结论·给三种人的三句话

### 一件必须先说清楚的事

第三编讲完了三个循环。它们看起来毫不相干，实际上是同一张脸：都在短期指标上表现为改善，失效的时候都不报警，而且都会自我加固。

还有第四个共同点，也是最不好受的一点：

**这三个循环里，没有任何一个是因为有人做错了什么。**

出题的人在提供支架，老师在及时反馈，孩子在参考同伴，学校在追求效率，平台在细化知识图谱，家长在关心进度。每一个局部都在解决一个真实的问题，而且解决得不错。

可这些局部的好加起来，结果是学习所需要的那点东西被挤掉了。

这就带来一个很实际的困难：**当没有人做错事的时候，改进从哪里开始？**

我见过两种回答，都不好。

第一种是把责任推给个人：老师要有情怀，家长要有耐心，孩子要有内驱力。这种话说起来痛快，落到地上什么也不是——**因为它要求每个人在自己的约束之外行动。**

第二种是等待系统改变：等评价改革，等课时松动，等有人把整件事重新设计一遍。这种话听起来清醒，实际上是一种体面的放弃。

我的回答是第三种，而且很朴素：

**每个人都在自己真正有权限的那一处，留一个口子。**

不是改造系统，是留一个口子。口子不大，但它必须是真的——真的能让一个还没算清楚的差异，活到下一轮。

下面三句话，就是三种人各自能守住的那个口子。

### 给老师的一句话

你能守住的口子，在**时机**上。

这本书里所有的方法——五根柱子、四轮改写、五步校航、一正一负——如果你只能记住一件事，请记住这个：

**让孩子先说，你再说。**

不是让你少说。你该讲的照讲，该纠正的照纠正，该给的支架照给。**只是把顺序调过来：在他形成自己的判断之前，别给方向性的线索。**

这一件事的成本几乎为零。它不需要多一分钟课时，不需要换教材，不需要说服任何人。它只需要你在那关键的十秒钟里，忍住不说那句已经到嘴边的“想想我们昨天讲的方法”。

十秒钟。这就是全书写了五万字要争取的东西。

而它能同时对付三个循环：孩子先说，落差就属于他自己（第七章）；孩子先说，那个差异就有了活下去的最短时长（第八章）；孩子先说，他就必须自己决定这道题跟什么有关，而不是先去找箱子（第九章）。

如果你还有余力，加第二件：**每节课，只挑一处，问一句“原来那个方法从哪一步开始不够用了”。**

一处就够。不要每道题都问——那会变成新的形式主义，而形式主义是这本书从头到尾的敌人。

## 给家长的一句话

你能守住的口子，在**问题**上。

孩子放学回来，我们最容易问的是三句话：今天考得怎么样？这道题会不会做？作业写完了吗？

这三句都指向同一个东西：**结果**。而结果，正是这本书从第一页起就在说的那件不够用的证据。

如果你愿意换一句，我建议这一句：

**"这道题，有没有哪个地方跟你原来想的不一样？"**

这句话的好处是它不检查、不评判、不施压。它问的是**落差**，而落差是学习真正发生的地方。

孩子答"没有"，也没关系——多数时候他会答没有。但这句话问上一个月，他会开始留意"我原来是怎么想的"。这个留意，就是第七章说的"预测先行"。

有两件事请不要做，它们比不问更糟：

**第一，不要在他还没想完的时候提示。**家里的辅导比课堂更容易犯这个毛病，因为一对一的时候，看着孩子卡住实在太难受了。但那份难受是你的，不是他的。**他卡住的那两分钟，是这本书里最贵的两分钟。**

**第二，不要把"我孩子会好几种解法"当作骄傲的理由。**第一章说过：五种公式套法，不如清楚知道"总量固定"变成"单位量固定"时，比例关系为什么会反过来。**方法的数量是最容易看见、也最不可靠的指标。**

## 给正在用 AI 学习的孩子的一句话

这句话我想直接说给你听。

你现在拥有一样我们这代人上学时没有的东西：任何一道题，你都可以在几秒钟内拿到解法、步骤和讲解，而且讲得比大多数人都清楚。

这是一件好事。**我不打算劝你少用它，那种劝告既没用也不诚实。**

我只想让你知道一件事：**这样东西会不会帮到你，几乎完全取决于你在哪一秒打开它。**

题目看完就打开——你会得到一个正确答案，和一次什么也没发生的学习。

自己先写下一句"我觉得这题应该是……因为……"再打开——同一个工具，同一段回答，作用完全反过来了。

差别不在它，在那句话。

所以这句话是：

**在你问它之前，先写下你自己的猜测和理由——哪怕是错的，哪怕只有一行。**

错的猜测比没有猜测值钱得多。因为只有先有一个猜测，AI 的回答才能在你这里造成一次落差；没有猜测，它的回答只是流过去了。

还有一个用法，是这一代人才有的特权。看完 AI 的解答之后，再追问一句：

**"这道题还能被归到哪些别的地方去？分别是因为什么关系？"**

这个问题它答得极好，而且它是第九章说的那个拆箱动作。你手上这台机器，既是最擅长造箱子的，也是最擅长拆箱子的——**取决于你问它哪一句。**

## 最后，写给那些能改结构的人

如果你是教研员、校长、教材编者、平台设计者——你能守住的口子比前面三种人都大，所以我说得直接一点。

**第一，给“没有结算完的东西”留一个字段。**

现在的记录系统里，一切都是完成态：完成率、掌握度、正确率、达标人数。没有任何一栏叫“本周留下的真问题”。

**只要有一栏，它就存在；没有那一栏，它在组织里等于没有发生过。**这一栏不需要计分，不需要排名，只需要存在，并且在下一轮被读一次。

**第二，在任何一次提效之前，先规定省下来的时间归谁。**

新工具、新流程、新平台，都会省下时间。如果不预先规定，省下来的时间会被自动填满——填的一定是更多同类练习，因为那是最容易产生记录的东西。

**规定它的去处，是一个纯行政动作，成本为零，效果比任何教学倡导都大。**

**第三，给分类留一条回写的通道。**

孩子和老师发现的跨章节关系，现在没有地方去。留一个入口，一学期整理一次，看看有没有一条值得写进下一轮教研材料的。

**你不必接受所有的改类请求——那不现实。**但一个从来不接受任何改类请求的系统，会在数据越来越漂亮的同时，离真实的学习越来越远。

## 最后一句

这本书从一张打了勾的卷子开始，到这里可以合上了。

我知道它要求的东西，在今天的条件下不算容易：让孩子先说，等那十秒，留那一栏，问那一句。

但这些都不是理想主义。它们都是很小的、具体的、明天就能做的事。**这本书的全部主张，就是这些小事叠加起来的效果，比任何一次宏大的改革都更接近学习本身。**

回到开头那句话。它现在应该有了完整的分量：

**真正的会，不是从此不再迷路。真正的会，是旧地图失效之后，还能重新长出路来。**

而我们能为一个孩子做的，说到底也很简单：

**在他还能长的时候，别把地面铺死。**

## 后记·这本书是怎么写出来的

写完这本书，我想把它的做法交代清楚。不是为了自证，是因为**做法本身也是这本书的一部分**——如果一个人主张“要看结构怎么形成”，那他自己的书是怎么形成的，就不该藏着。

### 为什么去请那些不相干的学问

一开始我并没有打算写九章。我只想回答一个问题：**两个孩子都做对了，凭什么判断哪一份正确里面还留着东西。**

在教育学区内部找了很久，找不到一个能落到作业本上的答案。理解、迁移、深度、素养——每一个都对，每一个都用不上。后来我意识到问题出在哪里：**这些概念全都从“已经发生的行为”里读出学习。**做对了读出会，做错了读出不会。而我要问的恰恰是那个没有发生出来的部分。

于是我换了个方向：**去问那些不研究人、却认真研究过“什么才算真的存在”的学问。**

方法是固定的：每一章找三门彼此不相容的学问，让它们对同一个问题各自回答，然后不去调和，而是追问——**三家为什么不能温和地相加？它们共同默认了什么？**最后从三家自己的材料里，找出推翻那条共同默认的东西。

这个做法有一条硬检验，我在每一章都执行了：**把其中一门学问整条删掉，这一章的判断还站得住吗？**站得住，那门学问就只是修辞，应该删掉。九章二十七学科入口，都过了这一关。

我要感谢王德生博士。2019年起我跟随他研修，这套“让不相容的东西撞出第三样东西”的方法，来自他。

更具体的一笔账也要记清楚：这本书的地基，来自他的《课堂学习发生学》。“**课堂每天都在生产它看不见的东西，并把看不见的部分算作零**”——这一句是我这本书里所有“越好越糟”的判断的源头；“**账本每完备一分，那一段就被压掉一分**”是第三编三章的共同形状；“**它无人持有，因而不能在任何一个时点对任何一个人结清**”则解释了为什么这本书所有的分界都不肯落在当堂。前言第二节和理论章第一之二节，是我对这笔账的正式记录。

这本书里所有走得通的地方，方法上的功劳是他的；走不通的地方，是我的。

### 这本书和上一本

《课堂的智慧》处理的是一般课堂，是一份写给愿意跟着论证走完全程的读者的理论纲领。这一本把同一副眼光收进数学这一门具体的学科，也收进一种可以直接用的形式。

**上一本回答“课堂为什么会这样”，这一本回答“那么这道题该怎么问”。**两本可以分开读。

### 关于人工智能的参与

这本书的资料检索、跨学科材料的整理、初稿的文字组织和排版，有人工智能参与。全书的问题、结构、九组学科的选择、每一处判断与取舍，由我负责。

我把这件事写出来，是因为这本书有一整章在讲**过早的正确如何替代学习**。如果我自己用一个能瞬间给出流畅正确的工具，却假装它不存在，那这本书的第七章就先失效了。

**真正要紧的不是用不用它，是它在哪一秒进来。**这本书里每一处判断，都是我先写下自己的看法，再去查、去撞、去改的。这个顺序，正是我在第七章请每一位读者守住的那个顺序。

## 这本书还欠着什么

前言的最后一节写了三个请求，其中两个其实是欠条：**案例大多是构想的典型场面，需要真实课堂来替换；第七章那个数字来自英国的题库，需要中文题库重算一次。**这里补上第三条，它比那两条更难还。

### 延迟回生的证据。

全书判据的核心是“隔两三周再来一次”——换掉表示方式，看那条关系还能不能重新长出来。而**没有任何一个公开的数据集跨越这个时间窗**并且保留了结构变式任务。学习平台的日志记录得极细，但它记的是当天的每一次点击；标准化测验隔得够久，但它测的是同一批题型。

**这一条只能靠课堂本身。**它需要一位老师愿意在一个学期里，把同一条关系用三种表示方式、隔三周各问一次，并且把学生说不出来的那些时刻也记下来。

三件事都不影响你从现在开始用这本书。但它们决定了第二版能不能比第一版更硬。

## 最后

如果这本书只留下一句话，我希望是这一句——它不需要任何术语，任何人都能在明天早上用：

**让孩子先说，你再说。**

阳涌

## 参考文献

按章分列。同一文献在多章使用时，只在首次出现的章下列出。

### 本书直接承接的两部专著

1. 王德生 (2026)。《课堂学习发生学——一节课留下的那样东西，账本上没有它的字段》。德麦国际出版社，258 页，ISBN 978-1-970820-90-4。（本书前言第二节与理论章第一之二节的全部理论来源）
2. 阳涌 (2026)。《课堂的智慧——课堂内生对象、运行条件与失效动力的可证伪理论纲领》。德麦国际出版社，262 页，ISBN 978-1-970820-95-9。（本书理论章其余部分的来源；本书为其数学篇与读者版）

### 导论与第一章

3. Barbier, F. F., Dun, E. A., Kerr, S. C., Chabikwa, T. G., & Beveridge, C. A. (2019). An update on the signals controlling shoot branching. *Trends in Plant Science*, 24(3), 220–236.
4. Wang, Y., & Jiao, Y. (2018). Axillary meristem initiation—a way to branch out. *Current Opinion in Plant Biology*, 41, 61–66.
5. Wang, B., Smith, S. M., & Li, J. (2018). Genetic regulation of shoot architecture. *Annual Review of Plant Biology*, 69, 437–468.
6. Luo, Z., Janssen, B. J., & Snowden, K. C. (2021). The molecular and genetic regulation of shoot branching. *Plant Physiology*, 187(3), 1033–1044.
7. Bamforth, D. B., & Finlay, N. (2008). Introduction: Archaeological approaches to lithic production skill and craft learning. *Journal of Archaeological Method and Theory*, 15(1), 1–27.
8. Lew-Levy, S., Milks, A., Lavi, N., Pope, S. M., & Friesem, D. E. (2020). Where innovations flourish: An ethnographic and archaeological overview of hunter–gatherer learning contexts. *Evolutionary Human Sciences*, 2, e31.
9. Pargeter, J., et al. (2023). Replicability in lithic analysis. *American Antiquity*, 88(2), 163–186.
10. National Gallery, London. *Pentimento; The Madonna of the Pinks: Underdrawing and Changes of Mind*. 技术研究报告。
11. Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
12. Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008–1022.
13. Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child Development and Education in Japan* (pp. 262–272). W. H. Freeman.
14. Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology*, 82(3), 436–455.
15. Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36(5), 404–411.
16. Watson, A., & Mason, J. (2005). *Mathematics as a Constructive Activity: Learners Generating Examples*. Lawrence Erlbaum Associates.

17. Mason, J. (2002). *Researching Your Own Practice: The Discipline of Noticing*. RoutledgeFalmer.
18. Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66(2), 211–227.
19. Marton, F., & Booth, S. (1997). *Learning and Awareness*. Lawrence Erlbaum Associates.
20. Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100.

## 第二章

21. Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of Learning from a Modern Multidisciplinary Perspective* (pp. 1–51). Information Age.
22. Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199.
23. Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
24. Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989–998.
25. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
26. Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.
27. Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239–264.
28. Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed Cognitions* (pp. 47–87). Cambridge University Press.
29. Black, H. S. (1934). Stabilized feedback amplifiers. *Bell System Technical Journal*, 13(1), 1–18.
30. NASA. *NASA Systems Engineering Handbook* (SP-2016-6105 Rev 2). 关于验证与确认的区分。
31. Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction*, 18(6), 565–579.
32. Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13(2), 145–182.
33. Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65–97). Macmillan.

### 第三章

34. Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), 663–676.
35. Holmberg, J., & Perlmann, T. (2012). Maintaining differentiated cellular identity. *Nature Reviews Genetics*, 13(6), 429–439.
36. Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2), 251–267.
37. Koselleck, R. (2018). *Sediments of Time: On Possible Histories*. Stanford University Press.
38. Zou, Z., et al. (2024). Dynamic covalent chemistry and covalent adaptable networks. *Chemical Reviews*. 关于动态共价网络的综述。
39. Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22, 1–36.
40. Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12, 151–169.
41. Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press.
42. Runesson, U. (2005). Beyond discourse and interaction: Variation, a critical aspect for teaching and learning mathematics. *Cambridge Journal of Education*, 35(1), 69–87.
43. Fiorella, L., & Mayer, R. E. (2015). *Learning as a Generative Activity*. Cambridge University Press.
44. Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, 24, 335–359.

### 第四章

45. Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
46. Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711–733.
47. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (1996). Niche construction. *The American Naturalist*, 147(4), 641–648.
48. Laland, K. N., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology*, 30, 191–202.
49. Bordass, B., & Leaman, A. (2005). Making feedback and post-occupancy evaluation routine 1: A portfolio of feedback techniques. *Building Research & Information*, 33(4), 347–352.
50. Way, M., & Bordass, B. (2005). Making feedback and post-occupancy evaluation routine 2: Soft landings. *Building Research & Information*, 33(4), 353–360.
51. Göçer, Ö., Hua, Y., & Göçer, K. (2015). Completing the missing link in building design process: Enhancing post-occupancy evaluation method. *Building and Environment*, 89, 14–27.

52. Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books.
53. Papert, S., & Harel, I. (1991). Situating constructionism. In I. Harel & S. Papert (Eds.), *Constructionism*. Ablex.
54. Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, 3(3), 265–283.
55. Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
56. Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

## 第五章

57. Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical Problem Solving: An Analysis of Clinical Reasoning*. Harvard University Press.
58. Bowen, J. L. (2006). Educational strategies to promote clinical diagnostic reasoning. *New England Journal of Medicine*, 355(21), 2217–2225.
59. Schmidt, H. G., & Rikers, R. M. J. P. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education*, 41(12), 1133–1139.
60. McQuade, C. N., et al. (2024). Characteristics differentiating problem representation synthesis between novices and experts. *Journal of Hospital Medicine*, 19(6), 468–474.
61. Prakken, H. (2004). Analysing reasoning about evidence with formal models of argumentation. *Law, Probability and Risk*, 3(1), 33–50.
62. Allen, R. J. (2014). Burdens of proof. *Law, Probability and Risk*, 13(3–4), 195–219.
63. Kim, C. (2014). Adversarial and inquisitorial procedures with information acquisition. *The Journal of Law, Economics, and Organization*, 30(4), 767–803.
64. Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 195–215.
65. Kalia, A. A., Schrater, P. R., & Legge, G. E. (2013). Combining path integration and remembered landmarks when navigating without vision. *PLoS ONE*, 8(9), e72170.
66. Pólya, G. (1945). *How to Solve It*. Princeton University Press.
67. Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.

## 第六章

68. Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *The Quarterly Journal of Experimental Psychology*, 28(2), 255–270.
69. Jones, M. R., Summerell, L., & Marshburn, E. (1987). Recognizing melodies: A dynamic interpretation. *Quarterly Journal of Experimental Psychology*, 39A(1), 89–121.
70. Peretz, I. (1987). Shifting ear differences in melody comparison through transposition. *Cortex*, 23(2), 317–323.

71. National Institute of Standards and Technology. (2015). *FIPS PUB 180-4: Secure Hash Standard*.
72. National Institute of Standards and Technology. (2023). *FIPS PUB 186-5: Digital Signature Standard*.
73. Goldwasser, S., Micali, S., & Rackoff, C. (1989). The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1), 186–208.
74. Goodnow, C. C., Adelstein, S., & Basten, A. (1990). The need for central and peripheral tolerance in the B cell repertoire. *Science*, 248(4961), 1373–1379.
75. Matzinger, P. (1994). Tolerance, danger, and the extended family. *Annual Review of Immunology*, 12, 991–1045.
76. Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A Study of Thinking*. Wiley.
77. Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, 19(6), 585–592.
78. Robinson, J., Chuang, C.-Y., Sra, S., & Jegelka, S. (2021). Contrastive learning with hard negative samples. *ICLR*.
79. Kullberg, A., Runesson Kempe, U., & Marton, F. (2017). What is made possible to learn when using the variation theory of learning in teaching mathematics? *ZDM Mathematics Education*, 49, 559–569.

## 第七章

80. Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
81. Yagishita, S., et al. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science*, 345(6204), 1616–1620.
82. Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026.
83. Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *American Economic Review*, 87(5), 847–862.
84. Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
85. Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 677–691.
86. Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96–146.
87. Wang, Z., Lamb, A., Saveliev, E., Cameron, P., Zaykov, Y., Hernández-Lobato, J. M., Turner, R. E., Baraniuk, R. G., Barton, C., Peyton Jones, S., Woodhead, S., & Zhang, C. (2021). Results and insights from Diagnostic Questions: The NeurIPS 2020 Education Challenge. *Proceedings of Machine Learning Research*, 133, 191–205.
88. Eedi. (2024). *Eedi — Mining Misconceptions in Mathematics*. Kaggle 竞赛公开数据集。本书第七章的全部读数由该数据集的公开标注子集计算而得。

## 第八章

89. Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(4), 902–904.
90. Whitt, W. (1992). Understanding the efficiency of multi-server service systems. *Management Science*, 38(5), 708–723.
91. McNamara, J. M., & Houston, A. I. (2003). Assessing predation risk: Optimal behaviour and rules of thumb. *Theoretical Population Biology*, 64(4), 417–430.
92. Crane, A. L., et al. (2022). Exploratory decisions of Trinidadian guppies when uncertain about predation risk. *Animal Cognition*, 25(3), 581–587.
93. Ioannou, C. C., Ramnarine, I. W., & Torney, C. J. (2017). High-predation habitats affect the social dynamics of collective exploration in a shoaling fish. *Science Advances*, 3(5), e1602682.
94. Kochetova-Kozloski, N., & Messier, W. F. Jr. (2010). The effect of benchmarked performance measures and strategic analysis on auditors' risk assessments and mental models. *Accounting, Organizations and Society*, 35(3), 316–333.
95. Courty, P., & Marschke, G. (2004). An empirical investigation of gaming responses to explicit performance incentives. *Journal of Labor Economics*, 22(1), 23–56.
96. Woelert, P. (2021). Reactivity and the dialectics of performance measurement. *Administration & Society*, 53(6).
97. Hollnagel, E. (2009). *The ETTO Principle: Efficiency-Thoroughness Trade-Off*. Ashgate.
98. Cyert, R. M., & March, J. G. (1963). *A Behavioral Theory of the Firm*. Prentice-Hall.
99. Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
100. Repenning, N. P., & Serman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened. *California Management Review*, 43(4), 64–88.
101. Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1), 67–90.

## 第九章

102. International Commission on Zoological Nomenclature. (1999). *International Code of Zoological Nomenclature* (4th ed.).
103. de Queiroz, K. (2022). Systematics and the Darwinian revolution. *Philosophy of Science*, 89(2), 246–266.
104. Podani, J. (2009). Taxonomy versus evolution. *Taxon*, 58(4), 1049–1053.
105. International Organization for Standardization. (2020). *ISO 668:2020, Series 1 Freight Containers — Classification, Dimensions and Ratings*.
106. Levinson, M. (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton University Press.
107. Egyedi, T. M. (2000). The standardised container: Gateway technologies in cargo transport. *SIIT Conference Proceedings*.
108. Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

109. Bernstein, B. (1971). On the classification and framing of educational knowledge. In M. F. D. Young (Ed.), *Knowledge and Control* (pp. 47–69). Collier-Macmillan.
110. Pountney, R., & McPhail, G. (2021). Curriculum integration and the semicentennial of Basil Bernstein's classification and framing of educational knowledge. *British Journal of Sociology of Education*, 42(5–6), 715–730.
111. Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press.
112. National Research Council. (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8*. National Academies Press.
113. Whitehead, A. N. (1929). *The Aims of Education and Other Essays*. Macmillan.
114. Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How People Learn: Brain, Mind, Experience, and School*. National Academies Press.

## 附录一·九张操作卡

这九页可以单独印出来，贴在教案本或者备课夹里。每一张对应正文的一章。

用法只有一种：自己看，用来决定下一句话怎么问。不要用来打分，不要用来排名。

### 卡一·没走的那条路还在不在

**这一章判断的：**做对之后，那些没走的路还在不在。

**危险格：**封死的正确——答案、步骤、分数都很好，可原来那条路之外什么也没有。

**三个动作**

1. **换掉一个关键条件，不是换数字。**把“每人分得同样多”改成“每组人数不同”；把等腰改成一般三角形；把等可能改成不等可能；把唯一解的方程改成恒等式或矛盾式。
2. **问一句失效点：**“原来那个方法，从哪一步开始不能往下走了？”
3. **隔两三周再来一次，**换一种表示方式，触发同一个结构。

**记录五栏（只记事实）**

对象重定了吗·单位重定了吗·关系重建了吗·失效点找到了吗·枝子回生了吗

**一句提醒：**余枝的价值不由数量决定。五种公式套法，不如清楚知道“总量固定”改成“单位量固定”时，比例关系为什么反过来。

### 卡二·这个答案，是谁在替他撑着

**这一章判断的：**撑着这次正确的是哪根柱子。

**危险格：**借来的正确——答案对，可撑着它的是题面、模板和提示。

**五根柱子**

对象·单位·前提·表示·支架

**四个动作**

1. **设计支架扰动，不是加难度。**改题之前先问：我改的是计算量，还是承重结构？
2. **答案之外让他写三句话：**原来那个方法靠什么成立？这次的变化让哪一步不再够用？现在由什么接手？
3. **区分“用了提示”和“依赖提示”**——提示撤掉以后，那部分工作有没有去处。
4. **老师给自己画一张载荷图：**这节课里，孩子的哪一部分推理其实是我替他做的？

**一句提醒：**撤支架是设计动作，不是诊断手段。有些工具（语言辅助、无障碍工具）不是脚手架，是接口，撤掉它不叫暴露真实能力。

### 卡三·旧知识会不会被改写

**这一章判断的：**学了新的，旧的动了没有。

**危险格：**标本式的正确——新旧各占一格，各自都对，互不相干。

**四轮设计**

1. **原型轮：**在熟悉材料里把关系稳定建立起来。这一轮不能省。
2. **单替换轮：**只换一样东西。换两样以上，他分不清是哪一处让旧方法失效的。
3. **重合成轮：**用新对象重建，并说出哪一条约束必须保住。

4. **回写轮**：回到原型，问那一句——“我们上次说的那句话，现在还完全对吗？哪里要改？”

没有第四轮，前三轮只是变式练习。

#### 记录四栏

拆出了什么·保住了什么·换了什么·改写了什么

第四栏空着，就说明这一轮没有代谢，只有正确。

**用 AI 时加三道口子**：问之前先写旧方法可能在哪儿失效；拿到多种方案后删掉一种破坏了约束的；关掉之后说出哪个词、哪条边界因为这次对话被改写了。

## 卡四·留下下一堂课绕不开的东西

**这一章判断的**：下一堂课能不能绕开这一堂。

**危险格**：蒸发的正确——热闹、准确、被表扬，下一节课完全可以绕开。

#### 备课前六问

1. 谁会使用孩子的数学表达，并据此继续做什么？
2. 原来的表达，在哪一点刚好不够用了？
3. **孩子产生的哪一处差异，将被写进下一轮的任务、工具或规则？**
4. 他们要建成什么东西，是作者不在场也能运行的？
5. 陌生的使用者会在哪一步误用、停住，或者产生不同的解释？
6. **这次故障，具体改变下一轮的哪一个条件？**

第三问和第六问是骨头。长期答不上来，无论课上得多好，都是蒸发的。

#### 插手点对照

| 阶段   | 该做的           | 最容易做错的     |
|------|---------------|------------|
| 交付   | 制造信息差         | 预先给出唯一标准表达 |
| 凝结   | 追问哪处差异有区分力    | 把所有差异一次讲完  |
| 改写环境 | 保存痕迹，重排后续任务   | 让成果停在展示    |
| 造物   | 提供真实使用者       | 站在作者旁边替他补漏 |
| 故障   | 区分是使用者还是造物的问题 | 只给分数或笼统评价  |
| 回写   | 把故障写进下一轮条件    | 把反馈封存在本节课  |

## 卡五·第一次走偏的时候

**这一章判断的**：方向是谁给的。

**危险格**：导航式的正确——一路没错，撤掉提示就不知道在哪儿。

#### 出发前三问

1. 至少哪两个诊断，会给出不同的下一步？
2. 我暂时选哪一个？什么证据让它领先？
3. **途中应该看到什么？看到什么我必须改判？**

#### 走偏后一问

**这次要改的，是位置、是地图，还是航标？**

算错一个数 → 改位置。把迫及当成相遇 → 改地图。一路自我感觉良好直到最后才发现不对 → 航标设错了，下次换检查点。

### 三条提醒

别做稻草人（两个候选里至少一个来自孩子真实的倾向）· 两个候选就够（标准是会导致不同的下一步，不是多）· 偏差出现时先别急着纠正，给他十秒。

## 卡六·对“几乎一样”说不

**这一章判断的：**该不用的时候忍不忍得住。

**危险格：**过界的熟练——什么都能套上去，而且套得很顺。注意：它出现在“用得  
上”那一系列，通常是班上最好的那批孩子。

### 备课时准备一对

一个远的正例，一个近的负例。

近的负例要只破坏一个承重点，而且错误调用要产生看得见的后果。

### 动手前三问

1. 你最先想调用的是什么？（要在他动笔之前问）
2. 签名里的哪一项没有通过？
3. 所以现在改用什么，或者暂时还不能决定什么？

### 六对现成的

正比例 ← 起步价·平均数 ← 两班人数不同·等式变形 ← 两边同乘  $(x-1)$ ·相似全等 ←  
两角相等·线性 ← 局部像直线·独立 ← “不会同时发生”

**一句提醒：**“这道题我还不能判断，因为不知道有没有起始量”——这句话应该得分。

## 卡七·正确来得太早

**这一章判断的：**这道题分不分得出人。

**危险格：**安静的正确——全班一致，没有人再有疑问。

### 备课时一问

这道题，能让目标规则和哪一种常见错误规则给出不同的答案？如果两种规则仍然给出相同答案，这道题只能用来练习，不能用来诊断。

### 课堂上一守

方向性线索出现在孩子承诺之前，还是之后？

（情绪支持随时给；任务澄清随时给；“想想昨天那个方法”要看时机。）

### 几乎零成本的一个改动

全班同时写在纸上，再一起亮出来——而不是一个一个举手。一节课的独立判断，  
从 1 份变回 40 份。

### 三个读数

模型分辨率（有几道题有区分力）· 预测先行率（多少次是孩子先说）· 私人判断保留率（多少人在看见别人答案之前给出了答案）

## 卡八·课堂越高效，越难发生

**这一章判断的：**一个没算清楚的差异，能不能活到下一轮。

**危险格：**高效空心——快、准、齐，什么也没长出来。

### 三个读数

余量存活时长（差异从出现到被处理或放弃，隔了多久）·差异回写率（有多少具体改变了下一节课的某个条件）·不可审计证据的转化率（有多少进了某种正式记录，哪怕只是教案边上一行字）

### 两个动作

1. **每节课留一处口子**，并且给它一个具体去处：一个要回答的分歧、一个要设计的反例、一条要写进下一轮的规则。留空白而没有去处，孩子只会拿去发呆。
2. **每次提效之前，先规定省下来的时间归谁**。不规定，它一定会被同类练习填满。

**一句提醒：**一个老师在满载的课表里选择压平差异，是他约束下的正确选择。这一章说的多数事情不在一节课的权限之内——能改结构的人应该去改结构。

## 卡九·越容易搬运，越难长出来

**这一章判断的：**箱子打不打得开。

**危险格：**封箱的正确——箱内全对，出了箱就不认识。

### 三个读数

拆箱重构率（给一道不带任何标签的题，他能不能自己判断它跟什么有关）·分类回写率（他提出的新分组，有多少真的改变了后续任务）·双向边界辨识率（“这道题为什么不属于那一类？”）

### 一个几乎零成本的检验

期末把不同章节的题混编，去掉标题。这份卷子和按章节考的卷子之间的分差，就是箱子替他做了多少事。

### AI 的一个用法

不要问“这是哪一类题”，改问：

“这道题还能被归到哪些别的地方去？分别是因为什么关系？”

**一句提醒：**孩子提出的新分组不天然正确。“这两道题都有3”不是一条关系。改类请求也要举证：共享的是哪一条关系，什么反例会推翻它。

## 如果九张只能留一张

留卡七背面那一句：

**让孩子先说，你再说。**

它成本为零，而且同时对付三个循环。

## 附录二·这些想法从哪儿来

——一张阅读地图。

这本书里的每一个判断，都站在别人已经走了很远的路上。这一篇是把那些路标一次列清楚——一是为了不掠美，二是为了让你知道，读完这本书之后可以往哪儿继续走。

我按章排列。每一条只写两件事：它回答的是哪个问题，以及这本书在哪里跟它分手。

### 第一章·没走的那条路还在不在

**变异理论 (Ference Marton 及其同事)**

它回答的问题是：一个人凭什么能“看见”一个概念？答案是——只有当某个方面发生变化、而其他方面保持不变时，这个方面才可能被辨别出来。一个概念周围那些“本来可以不是这样”的可能性，构成了学习得以发生的空间。

这是本章最近的一位邻居。它和我的分手处在于：变异理论关心教学要提供什么变化，我关心变化过后学习者身上还剩下什么——那些没被走的路，是不是还带着“我为什么没走它”的说明。

**建议先读：** Marton & Booth, 《Learning and Awareness》(1997)；以及 Kullberg、Runesson Kempe 与 Marton 关于变异理论在数学教学中的综述 (ZDM, 2017)。

**为未来学习做准备 (John Bransford & Daniel Schwartz)**

它回答的问题是：为什么我们的迁移测验总是低估学习？他们指出，传统的迁移测验把人关在一个“不许查资料、不许试错”的房间里，只测他能直接搬走什么；而真实的能力常常表现为在新环境里更快地重新学习。

这本书第一章的整个立场，可以看作这条线在课堂尺度上的延伸。分手处在于：他们用后来的学习速度作为证据，我想找一个在后来的成功出现之前就能看见的东西。

**建议先读：** Bransford & Schwartz, "Rethinking Transfer" (1999)；Schwartz、Bransford & Sears, "Efficiency and Innovation in Transfer" (2005)。

**适应性专长 (Hatano & Inagaki) / 程序灵活性 (Star, Rittle-Johnson) / 生产性失败 (Kapur) / 学习者生成例子 (Watson & Mason)**

这四条线分别从“专长的类型”“方法的灵活”“失败的价值”“例子的生成”回答同一件事：熟练之外还有什么。正文第一章的分界一节，逐条谈过它们。

### 第二章·这个答案，是谁在替他撑着

**帮助的两难 (Koedinger & Aleven)**

它回答的问题是：什么时候该给帮助，什么时候该收回？帮助太多，学习者不必自己产生推理；帮助太少，他会卡死或者形成错误。“什么时候给”本身就是一个可以研究的科学问题——这句话正是第二章的骨架。

**建议先读：** Koedinger & Aleven, "Exploring the Assistance Dilemma" (2007)。

**分布式智能 (Roy Pea) / “用工具的效应”与“工具带来的效应” (Gavriel Salomon)**

它们回答的问题是：一次成功，究竟该记在人身上还是记在人—工具系统身上？这是第二章“五根柱子”里第五根的直接来源。

**学习与表现的区分 (Soderstrom & Bjork) / 提取练习**

它们回答的问题是：当下的流畅，能不能作为学习的证据？答案是常常不能。这条线是第二章“心理学”一节的全部依据。

**建议先读：**Soderstrom & Bjork, "Learning Versus Performance" (2015)。

**认知负荷理论 (Sweller) / 支架 (Wood, Bruner & Ross)**

第二章的“载荷”和认知负荷理论里的“负荷”不是同一个东西：前者说的是一条推理由谁承担，后者说的是工作记忆的占用量。两者经常同时起作用，但可以分开测量。

### 第三章·旧知识会不会被改写

**反身抽象 (皮亚杰)**

它回答的问题是：新的认识从哪里来？皮亚杰的答案是——从对自己已有动作的重新组织中来，而这种重组会反过来改变原先那些动作的意义。第三章“回头改写旧的”这一步，是这条线在数学课堂上的具体化。

**表征重述 (Annette Karmiloff-Smith)**

它回答的问题是：为什么孩子已经“会做”之后，还会经历一次明显的重组？她的模型认为，知识会被反复重新表征——从只能使用，到可以被言说、被检查、被跨领域调用。这是第三章最近的一位邻居；分手处在于她关心表征格式的层级变化，我关心的是旧知识的边界和适用域有没有被改写。

**建议先读：**Karmiloff-Smith, 《Beyond Modularity》(1992)。

**概念转变 (Posner 等) / 知识碎片 (diSessa) / 过程与对象的双重性 (Sfard)**

三条线各自回答“旧的怎么变成新的”。第三章不主张旧概念被清除，而主张它被降级成一个有条件成立的特例——这一点和概念转变的“替换”图景不同。

**建议先读：**Sfard, "On the Dual Nature of Mathematical Conceptions" (1991)。

### 第四章·留下一堂课绕不开的东西

**使用本位的语言习得 (Tomasello; Bybee)：**规则从大量具体使用中组织出来，而不是先存在再被调用。

**生态位建构 (Odling-Smee, Laland, Feldman)：**生物不只适应环境，还改造环境，而且被改造的环境会被后代继承。

**使用后评价与“软着陆” (Bordass & Leaman; Way & Bordass)：**建筑的性能知识在交付之后才开始出现。

**建造主义 (Papert) / 知识建构共同体 (Scardamalia & Bereiter) / 情境学习 (Lave & Wenger)：**正文里逐条划过界，它们是这一章绕不开的四位前辈。

### 第五章·第一次走偏的时候

**临床推理与问题表征 (Elstein, Shulman & Sprafka; Bowen; 疾病脚本研究)：**专家不是从症状检索病名，而是先把叙述压缩成有区分力的表征。

**可撤销论证与举证责任 (Prakken; Allen; 对抗制的信息获取模型)：**结论来自支持与反对论证的比较，而不是一条永不修订的推导链。

**导航中的路径积分与地标 (Foo & Warren 等; Kalia, Schrater & Legge)：**两类线索在冲突时如何整合与切换。

**波利亚**（《怎样解题》）／**元认知监控**／**自我调节学习**：第五章的五步和它们最容易被混淆，正文第五章有逐条分界。**如果你只读一本，读波利亚**——它是这条路上所有人的起点。

## 第六章·对"几乎一样"说不

**移调旋律的识别**（Cuddy & Cohen; Jones 等）：表面全变而关系保持，人仍能认出同一个主题；实验里专门加入"保留轮廓的诱饵"，因为相似会制造误认。

**消息摘要与数字签名**（NIST 的相关标准）／**零知识证明**（Goldwasser、Micali & Rackoff）："相似不是证据"，以及"验证不必泄露全部生成过程"。

**免疫耐受**（Goodnow 等的中央与外周耐受研究；T 细胞的动力学校验研究）：不响应是一种学出来的积极结果。

**对比学习中的困难负例**（机器学习）：和第六章的"欺骗性近邻"是同一个直觉的两种实现——一个用来训练模型，一个用来训练人。

## 第七章·正确来得太早

**模型可识别性**（统计学的基本概念）／**认知诊断评估中的 Q 矩阵可识别性**

后面这条尤其值得一提：**认知诊断评估**这一支，早就在问"一套题目能不能唯一确定学生的技能掌握模式"。第七章那句"这道题只能练习不能诊断"，在这条线里有成熟得多的数学表达。想把这一章做成真正的测评工具，应该从这里读起。

**奖励预测误差**（Schultz、Dayan & Montague）：没有预期，就没有落差。

**信息级联**（Bikhchandani、Hirshleifer & Welch; Anderson & Holt 的实验室验证）：公开的一致不是多份独立证据。

**反馈干预的效果**（Kluger & DeNisi）：反馈平均有效，但相当一部分反馈会让表现变差——取决于它把注意力引向哪里。

## 第八章·课堂越高效，越难发生

**重交通下的排队**（Kingman）／**多服务系统的效率**（Whitt）：利用率、波动、等待，三者不可能同时优化。

**能力陷阱**（Repenning & Sterman）

这一条是第八章最近的邻居，也是最值得单独读的一篇：**组织越是靠加班和救火来达成短期指标，越没有时间做那些能减少未来救火的改进；而短期指标恰恰在变好**。把"组织"换成"课堂"，把"救火"换成"讲题"，第八章的循环几乎可以逐句对应。

**建议先读**：Repenning & Sterman, "Nobody Ever Gets Credit for Fixing Problems that Never Happened" (2001)。

**组织松弛**（Cyert & March）／**效率—彻底性权衡**（Hollnagel）／**审计社会与指标反应性**（Power; Courty & Marschke; Woelert）／**古德哈特与坎贝尔定律**：一旦一个指标成为目标，它就不再是一个好指标。

**心理安全感**（Edmondson）：为什么有的团队里人敢说"我不确定"。

## 第九章·越容易搬运，越难长出来

### 分类及其后果 (Bowker & Star)

第九章最近的邻居。它回答的问题是：分类系统一旦建成，会怎样悄悄地决定什么能被看见、什么会消失在缝隙里。"基础设施在正常工作的时候是隐形的"——这句话是第九章整章的种子。

**建议先读：**Bowker & Star, 《Sorting Things Out》(1999)。

**像国家一样观看 (James C. Scott)：**治理需要把复杂实践变成可记录、可比较的单位；简化让治理成为可能，也会抹掉支撑实际运行的地方性知识。

**分类与框架 (Basil Bernstein)：**课程的边界强弱和控制关系，是教育社会学处理这个问题的经典框架。如果你觉得第九章说的是常识，那是因为伯恩斯坦在五十年前已经把它说进了这个领域的空气里。

**集装箱的历史 (Marc Levinson, 《The Box》)：**标准箱怎样重造了港口、船舶和整个贸易网络。

### 如果只读三本

- Marton & Booth, 《Learning and Awareness》——关于"凭什么看得见"。
- 波利亚, 《怎样解题》——关于"走偏了怎么办", 至今没有被超越。
- Bowker & Star, 《Sorting Things Out》——关于"箱子在替我们做什么"。

三本合起来, 覆盖了这本书三编的骨架。剩下的, 是我在它们中间的那点连接工作。

## 附录三·给研究者：怎样验证这些判断

这一篇写给愿意把书里的说法真正做成研究的人：研究生、教研员、做课堂实证的老师、教育技术团队。

正文里我没有放证伪清单和统计条件——那些放在正文里会打断阅读。但它们必须存在，否则这本书就只是一套说得通的话。一个不能被验证的教育主张，无论多好听，都只是修辞。

下面按三编各给一组。每一组都写清：用什么设计、看什么读数、出现什么结果我就该认输。

### 一、第一编（什么才算真的学会）

**核心可检验命题：**对“对象重定—单位重定—关系重建—失效点定位”的前置编码，能预测后续在结构变式任务上的表现。

**最省钱的设计：**几乎所有学校都存着阶段测验、作业和错题本。在不知道学生后续成绩的情况下，请人对旧作业做盲编码，再对照几周后的结构变式任务。不需要新的干预，不需要额外课时。

#### 该认输的四种结果

1. 控制住先前的数学成绩、工作记忆、阅读理解和已掌握的方法数量之后，这套编码**不再有增量预测力**。
2. 所谓的余枝，**永远只在孩子已经做对新题之后**才被观察到，从来不在成功之前出现——那它只是对结果的事后描述。
3. 两种教法对照：一组学多种解法但不讨论适用条件，一组只比较少数解法却明确标出对象、单位、承重条件和失效点。**前者在变式题和延迟任务上稳定不差于后者，那“排除关系比数量更重要”这个判断就该推翻。**
4. 经过训练之后，不同的编码者**仍然无法对“失效点”“对象重定”形成可接受的一致**——那这套东西承担不了评价功能，只能留作解释性的说法。

**到期赌注：**2031年12月31日之前，应当出现至少一项预注册、跨不少于六个同质教学单元的研究，证明前置编码在控制上述变量后仍能显著预测延迟结构变式表现。以下情况不算命中：只用学生自我报告；只测换数字不换结构的近迁移题；样本跨越制度差异极大的地区；先看结果再调整编码标准；编码发生在学生已完成迁移任务之后。

### 二、第二编（一堂课该留下什么）

三章各有一条独立的可检验命题，可以分别做，也可以放进同一个学期的对照设计。

#### 第四章（终点换手）

**命题：**环境回写（前一轮产生的差异真的改变了后一轮的任务、工具或规则）会先于迁移成绩发生变化。

预测的顺序是：先出现学生主动引用过去工具与错误的频率上升；随后出现陌生使用者独立使用学生造物的成功率上升；**最后才是标准迁移测验和概念解释的变化。**

该认输：如果顺序反过来——迁移成绩先动而回写指标不动——那么这一章说的机制不是原因。

## 第五章（校航）

**命题：行动前写下改判条件的频率**，会先于偏差处理质量变化，再先于陌生题迁移变化。

最省钱的对照：同一位老师、两个平行班，只改一件事——一个班在动笔前写“看到什么我必须改判”，另一个班照常。

该认输：如果两个班在延迟结构变式上没有差异，或者写改判条件的班只是把时间花在了抄写上（可以通过改判条件的内容质量编码检出），这一章的处方无效。

## 第六章（校界）

**命题：同时训练远表面正例与高相似负例**，比同等数量的正例训练产生更准的边界。

读数分两个方向：**错误接受率**（该拒绝的题上错误调用了熟悉关系）和**错误拒绝率**（该用的时候不敢用）。这两个必须分开报，只报总正确率会把这一章想看的東西盖住。

该认输：如果加入近邻负例只降低了错误接受率，却把错误拒绝率抬高到抵消收益的程度，那“耐受”这条处方需要重新设计——或者说明它只适用于某个熟练度区间。

## 三、第三编（为什么它这么少发生）

第三编是三个机制假设，它们的检验方式和前两编不同：**要看的是事件发生的先后顺序，不是某个干预的效果**。所以需要课堂时间序列数据，至少一个学期。

### 第七章（正确性消音）预测的顺序

一轮循环里，事件应当按这个次序出现：

① 不同解释产生相同答案的题目比例上升 → ② 学生主动写下预测与失败条件的次数下降 → ③ 公开表达疑问的学生比例下降 → ④ 教师或系统给出方向性提示的时间进一步提前。

该认输：如果这四项之间**没有稳定的时间先后关系**，或者顺序在多个班级里各不相同且无法归入“从不同入口进入”的三种路径，那么它就不是一个循环，只是四个同时存在的现象。

### 第八章（余量清零）预测的顺序

同时记录三样东西：课程利用率、学生的探索风险行为、评价的可审计化程度。用时间序列判断**哪一项先动**。三个入口（进度先动／风险先动／评价先动）各自有不同的后续顺序，正文里已经写清。

该认输：如果三项之间只有同期相关而没有可辨的先后，或者“余量存活时长”与后续的结构变式表现无关，这一章的机制不成立。

### 第九章（知识集装箱）预测的顺序

① 学生按章节位置而非关系证据判断题目同类 → ② 练习围绕标签形成固定步骤 → ③ 平台和教师的记录按该标签聚合 → ④ 低表现被解释为“该知识点未掌握” → ⑤ 下一轮增加同标签训练，而不是修改边界。

该认输：如果在去掉标签的混编测验上，学生表现与带标签测验**没有显著差距**，那么“箱子替他做了很多事”这个判断就站不住。

**一个共同的认输条件**：如果在任何一项设计里，**高效率、高一致或高标准化的班级同时在结构变式上也更好**，那么第三编三章的核心判断整体错误——这三个循环不是代价，只是效率的正常伴随。

## 三之二、第七章那次真算：怎么复算

第七章的四个数字来自一份公开数据集的标注子集：**1869 道题、4370 个已标注误解的错误选项、1604 种误解、757 个知识点**。复算步骤如下，任何人都可以在一台笔记本上跑完。

**第一步，取数。**Eedi 2024 年公开的“数学误解挖掘”数据集，每一行是一个（题目 × 错误选项 → 误解编号）三元组，并附题目所属的主题与知识点名称。

**第二步，两个定义。**

- **单题可分辨率** = 这道题标出的误解数 ÷ 这个知识点上全部已记录的误解数。**不可识别率** = 1 - 它。
- **成对分辨率** = 在同一知识点上任取两种误解、任取一道题，这道题**同时**为这两种误解准备了错误选项的比例。这是最严格的口径：只有两人给出**不同的**错误答案，才算被这道题分开。

**第三步，读数。**加权口径下：单题可分辨率 44.5%，不可识别率 55.5%；限定在题目数 ≥ 5 的 101 个知识点上，不可识别率升到 73.9%；成对分辨率 7.1%（按知识点取平均为 27.8%）；52.7% 的误解在整个题库里只有一道题测得出。

**这份计算的三个已知偏差，必须一并说清楚：**

1. **成对分辨率被低估。**一个孩子的误解即使没有对应选项，他也可能碰巧选中别的错误选项而暴露自己。所以 7.1% 是下界。但分辨力的上限是硬的——三个错误选项，最多三种误解。
2. **“已记录的误解”是下界。**只统计了标注者标出来的那些，真实存在的误解更多。**这一偏差与结论同向：**真实的不可识别率只会更高。
3. **这是英国的一份题库。**中文教材的题型分布不同，数字不能直接搬。**要做的复算是：拿一套中文题库，按同样的两个定义重算一次。**这是这本书最值得别人接手去做的一件事。

**如果复算的结果相反**——比如在中文题库上单题可分辨率超过 80%，或者成对分辨率超过 50%——那么第七章第一阶段“不可识别”就不是常态，只是特例，本书的三阶段循环应当从第一步开始修改。

## 四、几条方法上的提醒

**第一，读数必须在正文、检验和赌注三处使用同一个定义、同一个阈值、同一套编码规则。**三处不一致，结果无法解释。

**第二，编码者要不知道分组。**这套编码的判断空间很大，不盲编几乎一定会向假设倾斜。

**第三，不利的结果必须报告。**只报有利结果的研究，在这套主张上等于没有做。

**第四，注意伦理边界。**不能为了制造预测误差而制造羞辱；不能为了观察“真实能力”而撤掉某些孩子赖以进入问题的辅助工具；这类记录若被用于分班、排名或不利评价，必须公开编码规则、保留原始作品，并给学生和教师复核的通道。

## 最后

如果上面这些条件把书里的核心说法推翻了，我们仍然会有收获。

**如果一般能力解释了一切**，那么力气应该花在改善先备知识和认知资源上，而不是寻找余枝。

**如果方法的数量已经足够**，那么教学设计应该转向更经济的策略库建设。

**如果语言记录无法预测行为**，那么评价应该减少口述解释，增加真实任务表现。

这些都是有用的结论。**一个能被推翻的主张，推翻它的那一天也是有产出的一天——**这正是我把这一篇放进书里的理由。

## 名词索引

按首字笔画顺序排列。页码为该词出现的主要位置。

一正一负 87-88、90、118  
三层 18-19、24-27、30、34、55  
不变量 44-45、47-48、57  
中央耐受 85  
举证责任 74、76-77、80、117、135  
五类凭据 18、21-22  
交叉反应 85-86  
休眠 19、34、36  
余枝 37、39-41、130、138、141  
余量 25、28、101-103、105-108、117、133  
使用后评价 66-67、135  
使用本位 64、67、135  
信息级联 94-95、97、136  
借来的正确 29、49、51、130  
关系签名 84、86  
关系航标 76  
冻结结构型 19  
分配性失败 28  
利用率 102、105-106、136、139  
剥片 34-36  
动态共价 55-56、125  
参照系 44、47  
变异理论 134  
古德哈特 104、106、136  
可审计 105-106、108、133、139  
可撤销 74、76、135  
可识别性 92-93、96、113、136  
可识读性 113-114  
合意困难 45  
同分异构体 55-56  
命名的稳定 114  
地方性知识 113-114、137  
地标 75-76、135

增益递增 54、56  
失效点 38-40、47、130、138  
套语 64-65、67  
安静的正确 29-30、98、132  
完备度 18  
导航式的正确 29、72、78、131  
封死的正确 29、37、49、130  
封箱的正确 29、115、133  
庇护 103-105  
延迟回生 122  
悬置连接 19  
承重条件 22、84、86、88、138  
拆箱 28、116-117、120、133  
指标反应性 106、136  
换位区间 18  
探索 27、41、103-106、139  
提取练习 45、47、134  
摘要 83-84、86、136  
操作链 34-36  
改判条件 80、139  
数字签名 84、86、136  
方向性线索 80、94、97、99-100、132  
构式 64、67  
标准接口 114  
标本式的正确 29、58、130  
模型分辨率 99、132  
欺骗性近邻 86、88-90、136  
正在发生 15、37-38、49、58、69、78  
死的正确 15、24、29、37、49、130  
波动 46-47、101-103、105、136  
活的正确 15、37、49、58、69、78  
生态位建构 65、67、135  
生态继承 65、67  
疾病脚本 73、135  
相乘不相加 19  
知识点 22、84、92、96、105、110-111  
确认 46-47、67、83-86、103-104、124

私人信号 97  
移调 82-83、86、136  
空窗 19  
终点换手 25、68、71、138  
耐受 16、83、85-87、90、136、139  
能力陷阱 136  
腋芽 16、34、36  
自建条件 18  
蒸发的正确 29-30、63、69、131  
表征重述 135  
表现与学习 27  
观测设计 93、97  
证明标准 74、76  
诱饵旋律 83、86  
负反馈 46-47  
账本 18-20、36、121、123  
路径依赖 54、56  
路径积分 75-76、135  
转录因子 53-54、56  
软着陆 66-67、135  
载荷转移 49-50  
过早闭合 73-74、76  
过界的熟练 29、82、88、132  
适用范围 41、44、47  
重编程 53-54、56、61  
鉴别诊断 73、76、80  
问题表征 72-73、76-77、80、135  
零知识证明 84、86、136  
靶格 19、23-24、27、29  
音程 82-83、86  
顶端优势 34、36  
预测误差 94、97、136、140  
验证 10、38、46-47、49、84-87、94  
高效空心 29、107、133  
Eedi 88、95、127、140

---

## 答案越来越便宜的时候， 一张打了勾的卷子还能证明什么？

---

两个孩子都做对了。答案一样对，步骤一样齐，字迹一样清楚。哪一份正确里面还留着能继续生长的东西，哪一份已经把所有没走的路封死？这本书提出一个不靠“深不深”“真不真”来判断的问法，并把它做成一张贯穿全书的格子。

### — 九章，二十七门学问

植物学、考古学、密码学、免疫学、排队论、审计学……全部从零讲起，不预设你懂它们。

### — 九张操作卡，可以复印

每一章都从一个课堂现场开始，到一张明天就能用的卡结束。

### — 三个不太好受的解释

为什么课堂越高效、知识越好搬运，学习反而越难发生。

适读：小学至初中数学教师、教研员，以及愿意认真陪孩子想一想的家长。

同著者：《课堂的智慧——课堂内生对象、运行条件与失效动力的可证伪理论纲领》

---

德麦国际出版社

Demai International Press

数学教育 / 教师用书

定价：US\$ 10.50



9 781970 820966

ISBN 978-1-970820-96-6

---