

设对

王德生 · 德麦国际 SDE 学派 · SDE教育学 · 2026年8月10日 · 约 22,038 字 · 21 条参考文献 · 三种阅读模式

摘要

关于「教育是什么」，现行三种读法各自自治而互不相容：从此刻做出来的成绩读它（体育学），从把他带到这里的那条路径读它（历史学），从他所处的基本结构读它（正义学）。三者共同假定了一条未被说出的前提：一次教育的功过是一笔可以被分配完的量，只要找对归属的那一维，就能分完。推翻它的材料来自正义学自己——帕菲特的非同一性问题指出，一项足够深的介入会改变将来存在的是谁；因此「这样做让他更好了」这句话在一部分场合没有比较对象，不是难比，是另一侧没有那个人。据此本文提出：教育不是能力的养成，不是脉络的延续，也不是位置的补偿，而是**对设对的一次设置**。设对指任何一次比较性判断被放在另一侧的那样东西；它有时是可指认的人，有时是统计构造，有时在替代方案里根本不存在，而三家的记录里都不为它设字段。本文给出一句不含程度词的判据：**这份记录上，与被他比较的那一个，是谁？** 读数不是比率而是一份缺项清单。材料方面，本文对九十五条现行游泳世界纪录做了一次回溯检验：二〇〇七年及以前零条、二〇〇八至二〇〇九两年六条、二〇一〇至二〇一一两年零条，带器材注记者存活中位十七年、其余两年（Fisher 双侧 $p=5.3\times 10^{-7}$ ）。三条不利结果按原样写入正文，其中一条直接改掉了本文的措辞。

关键词：设对；虚对；设对清单；非同一性；参照点；教育本体

Title The Set Counterpart: Every Verdict of "He Has Improved" Places Something on the Other Side, and That Something Never Enters the Ledger

Abstract Three self-consistent yet mutually incompatible readings of "what education is" are current: reading it off the performance produced now (sport science), off the path that brought him here (historiography), and off the basic structure he occupies (theories of justice). All three share an unstated premise: that the credit and blame of an education form a quantity that can be fully allocated, so that identifying the right dimension of attribution suffices to allocate it. The premise is overturned by material belonging to the theory of justice itself. Parfit's non-identity problem shows that a sufficiently deep intervention alters who will exist; hence "this made him better off" lacks a comparison object in a class of cases — not hard to compare, but no one on the other side. Accordingly, education is here redescribed: not the cultivation of ability, not the continuation of a lineage, not the compensation of position, but **the setting of a counterpart**. The set counterpart is whatever is placed on the other side of a comparative verdict: sometimes an identifiable person, sometimes a statistical construct, sometimes nothing that exists in the alternative. None of the three ledgers provides a field for it. The paper offers one modality-free test question — on this record, who is the one he is being compared with? — and a reading that is not a ratio but a list of missing entries. A retrospective test on 95 standing

world swimming records found **zero from 2007 or earlier, six from 2008–2009, and zero from 2010–2011**; flagged records have survived a median of 17 years against 2 for the rest (Fisher two-sided $p = 5.3 \times 10^{-7}$). Three adverse results are reported in full; one of them forced a rewording of the paper's central claim.

Keywords set counterpart; void counterpart; counterpart inventory; the other side of comparison; non-identity; baseline change

一 问题：一句谁都听得懂而无人追问的话

家长会上，班主任说：这孩子这学期进步很大。

没有人追问。这句话听上去完全清楚——分数上去了，名次前移了，作业交得齐了。它是一句关于这个孩子的话。

但把它拆开会发现，它其实是一句关于**两样东西**的话。「进步」不是一个可以单独成立的属性，它是一个二元关系：这一次比**某一次**好。那么那一次是哪一次？是他上学期的分数？是同班同学的平均？是这个年龄段的常模？是他本来会有的样子？

四个答案给出四种完全不同的判断，而这句话里一个也没有说。更要紧的是——**记录上也没有**。翻遍学籍、成绩单、评语、素质报告册，你能找到他每一次的分数，找不到任何一栏写着「这一次是与谁比的」。

这不是记录得潦草。分数是精确的、日期是精确的、签名是齐全的。缺的那一栏根本没有被设置过，因为没有人认为它需要被设置——比较的另一侧被当作不言自明。

通行的解释有三种，每一种都成立，而三种不能同时成立。

第一种说：不需要写，因为看成绩就够了。他一百米跑进十三秒，这就是事实；跟谁比是次要的，成绩本身会说话。这一种把根据放在**此刻做出来的那个结果**上。

第二种说：不需要写，因为要理解一个人得看他怎么来的。把他从他的家庭、他的学校、他这三年经历的事情里抽出来，跟任何人比都是错的。这一种把根据放在**那条路径**上。

第三种说：需要写，但要写的不是某个同学，是他所处的结构。两个孩子的差距多半不是他们本人造成的，追问「跟谁比」只会把结构性的差别记到个人头上。这一种把根据放在**位置**上。

三种解释各自都能把那句话解释掉，而三种互不相容。第一种要求另一侧无关紧要，第二种要求另一侧根本不可比，第三种要求另一侧必须换成结构而不是人。要说其中两种错了，你得先说出一个凌驾于三者之上的标准——而这个标准正是不存在的东西。

本文的起点是把这个尴尬当真，并且换一个问法。不问「他到底进步了没有」，问：

当我们说他进步了的时候，我们在另一侧放了什么？而那样东西，在记录上叫什么？

二 三种现成的读法各自站在哪里

要看清共有前提，先得把三种读法摆到各自的位置上，并且承认它们各自是对的。

第一种：从此刻做出来的那个结果读它。

体育学是这一路最彻底的代表，因为它有一样别的学问没有的东西：秒表。跑完就是跑完，游完就是游完，成绩在电子计时器上，不需要解释、不需要还原脉络、不需要考虑家庭背景。运动科学的方法论文献把这一条写得很直白：在评价一项训练干预时，**成绩本身就是效标**——不是某个中间指标的代理，就是那个要被解释的东西（Hopkins, Hawley & Burke, 1999）。

这一路的立场可以概括成一句话：一个东西是什么，看它此刻做出来的结果就够了。

而它有一处自己承认的松动，本文后面要用到：运动科学自己发展出了「最小有价值变化」这个概念——一次成绩变化要大到某个值才算数，而这个值由该项目成绩的自然波动决定，不由运动员的能力决定（Hopkins, 2004）。也就是说，成绩会不会说话，取决于另一样东西的散布程度。

第二种：从把它带到这里的那条路径读它。

历史学是这一路最有纪律的代表，而它的纪律恰恰是禁令。柯林伍德在一九四六年主张：历史知识不是对外部事件的观察，是历史学家在自己心中重演过去的思想；离开那个语境，事件就不是那个事件。斯金纳在一九六九年把这条纪律做成了一份禁令清单——他把用后来的范畴去读前人的做法称为一整套「神话学」，其中最常见的一种是把一个思想家没有、也不可能有的意图安到他头上。

这一路的立场是：一个东西是什么，看它经什么路径成为现在这样就够了；而把它从路径里抽出来与别的东西并排，是一种可以被指认的错误。

第三种：从它所处的条件结构读它。

正义学是这一路最系统的代表。罗尔斯在一九七一年开宗明义：正义的首要主题是**社会的基本结构**——各种主要制度分配基本权利义务、划定社会合作利益的方式。他给的理由很硬：基本结构的影响极其深远且从一开始就在场，人们生来处在不同的位置上，而这些位置上的期望差异不是他们自己造成的。由此推出那句著名的判断：天赋的分布在道德上是任意的，没有人配得上他的自然禀赋所带来的那份优势。

这一路的立场是：一个东西是什么，看它在什么结构里、被分配到什么位置就够了。

三家不能同时为真。

取那句「这孩子进步很大」做例子。第一路要求成绩本身足以裁定，因此另一侧是谁不影响结论。第二路要求任何比较都必须在同一语境内进行，因此把两个孩子并排本身就是错误。第三路要求把差距归给结构，因此比较的另一侧必须从「另一个孩子」换成「另一种位置」。三个「另一侧该放什么」的答案互相排斥。

更尖锐的一处：体育学每天在做的那件事，正是历史学诊断为病理的那件事。体育学的整个成就体系建立在跨时空的可比性上：今天的一分四十二秒可以与四十年前的成绩并排、可以进同一张榜。而历史学说这正是时代错置的标准形态——把一件事从它的条件里抽出来，与另一件条件完全不同的事并排，然后宣布谁更好。

同样的形状出现在正义学与体育学之间。体育学的荣誉体系要求成绩归于运动员本人；而罗尔斯说，天赋在道德上是任意的，一个人不配得上它带来的收益。**这两句话不能同时为真。**一枚金牌要么是他的，要么不完全是他的。

三 三家共同假定的那一条

把三家的争论压成一句话，形状是这样的：

三家争的是——一次教育的功过该归给谁：归给此刻的表现、归给那条路径，还是归给那个位置。

一旦写成这个形状，那条共有前提就浮出来了。三家争的是归属的去处，而三家共同假定：

一次教育的功过是一笔可以被分配完的量——总有一个归属，只要找对了那一维，就能分完。

用更贴近本文的说法：「他变好了多少」是一个可以被结算的差额，而结算所需要的另一侧，是给定的。

这条前提有一个可以当场检验的性质：把它念给三家听，三家都会说「这还用说吗」。没有一位教练会否认成绩的进步归于训练与运动员。没有一位历史学家会否认一段经历有它的来龙去脉。没有一位正义论者会否认利益与负担最终要落到某些人头上——罗尔斯的整个差别原则正是一套关于怎么分的规则。不引起争论，正是共有前提的标志。

它也解释了教育评价里那些长期打不完的官司为什么打不完。「增值还是绝对水平」「过程还是结果」「个人努力还是家庭背景」——三方各执一词，而没有一方问过：这笔账一定分得完吗？分得完的前提是另一侧有一个确定的东西，那个东西是谁？

要撼动这条前提，从外面搬一个理由来是没有用的。说「比较是相对的」，三家会各自退回自己的阵地：教练说秒表不相对，历史学家说语境不相对，正义论者说结构不相对。必须用三家之一自己已经掌握、已经发表、只是没往这个方向用的材料。

那件材料在正义学手里，而且已经放了四十多年。

四 推翻它的那件材料：非同一性

一九八四年，帕菲特在《理与人》第四部分处理了一个他称之为非同一性的问题。它的原初场合与教育无关，是关于我们对未来世代有没有义务。

问题的形状是这样的。设想一个社会面临两条政策路线：一条耗竭资源，两百年后的人生活得很糟；另一条节制，两百年后的人生活得很好。直觉说前一条错了，因为它伤害了未来的人。

帕菲特指出这个直觉说不通。政策不只改变未来的人过得怎么样，它改变了**谁会出生**——不同的政策带来不同的工作、不同的迁徙、不同的相遇，两百年后活着的是完全不同的一批人。于是在耗竭路线下出生的那个人，不能说自己被弄糟了：**如果当初选了另一条路，他根本不会存在。**他没有一个更好的自己可以被比较；另一侧不是一个过得更好的他，是没有他。

四十年来这个问题在道德哲学里被反复讨论，主流的处理是承认它确实使「以人为中心的」伤害概念在这类场合失效，并转向非以人为中心的价值判断。**没有人把它拿来回答「教育的功过记在谁名下」。**它在自己的领域里回答的是「我们对尚未存在的人有什么义务」。

把它搬进教育，那条共有前提就塌了，而且塌得比原来更彻底。

其一，教育性的介入常常是足够深的那一类。换一所学校、遇到一位老师、被分进一个班、在十四岁那年被说了一句话——这些不只改变一个人拥有什么，它们改变他成为谁：他的兴趣、他的自我描述、他后来遇见的人、他所属的那个圈子。到十年之后，站在那里的那个人与「若无此事」的那个人，共有的东西可能少于我们以为的那么多。

其二，「他因此变好了」这句话在这类场合没有主词。它要求另一侧站着同一个人的另一种状态。而如果另一侧站着的不是他，这句话比较的就是两个不同的人——那已经不是「变好」，是「换了一个」。

其三，这不是测量问题。这一点最要紧，也最容易被读漏。把评价做得再精细，也无法把另一侧那个不存在的人测出来——**不是难比，是没有可比的那一个。**

其四，也是最不客气的一处：三家争的是这笔账该分给谁，而这件材料说的是，在一部分场合**这笔账没有可分给的那一方。**总额守恒的假定因此失效——不是分不匀，是这笔账在结构上不封闭。

五 设对：一类原来不算一样东西的东西

把正义学交出来的这件材料搬进教育，得到的不是一个悖论，是一个可以逐条清点的对象。

任何一次比较性判断——「他进步了」「他达标了」「他比去年好」「这个班比那个班强」——都必须在另一侧放一样东西，否则这句话不成立。

本文把被放在另一侧的那样东西，称为**设对**。

「设」字是承重的：它不是被发现的，是被**设置**的。是评价制度、是评分规则、是一位老师的一次比较，把某样东西放到了另一侧。换一套规则，另一侧就换一样东西，而被评价的那个人一动不动。

三件事必须马上说清，否则这个词会被读成三个已有的东西。

其一，**设对不是「参照标准」**。标准是被写下来的那条线（六十分及格、跑进十四秒）；设对是被放在另一侧的那样东西，它可以是一条线，也可以是一个人、一群人、一个统计构造、或者什么都不是。说得更硬一点：**一条标准本身也有它的设对**——那条线是从某个人群的分布里定出来的，而定它的那个人群才是真正的另一侧。第十三节会说明，这正是「标准参照」这个概念最容易被读漏的地方。

其二，**设对不是「反事实」**。反事实是一个模态陈述（若当初不这样，会怎样）；设对是关于**这一次判断**的陈述：这一次，制度实际把什么放在了另一侧。反事实问的是世界会怎样，设对问的是记录上写了什么。两者可以完全脱钩——一次判断可以有一个明确的设对（同届考生的平均分）而没有任何反事实依据。

其三，**设对有三种，而且它们在记录上长得一模一样**。这是本文全部可操作性的来源，下一节展开。

于是「教育是什么」这个问题有了第四种答案：

教育不是能力的养成，不是脉络的延续，也不是位置的补偿，而是对设对的一次设置。

一次教育所做的，最稳定、最不可撤回的一件事，不是让某个人拥有了什么，而是**决定了他此后将被放在谁的旁边来读**。分班、分层、分校、分轨、发一张什么样的证书、进一个什么样的圈子——这些动作改变的首先不是他的能力，是他的设对。而设对一旦被设置，后面的一切读数都在它上面产生。

最要紧的一句在这里：**设对不是任何一个人的属性**。它是一次判断与它的另一侧之间的关系。你没法说「这个学生的设对」，就像你没法说「这个数的大于」。而现行的全部教育记录——学籍、成绩、评语、档案、评估报告——**没有一份为「这一次是与谁比的」设过字段**。

这就是本文要说的那件事：设对在三家的账本上不是记录得少，是**压根不算一样东西**。体育学只记成绩，历史学只记留下痕迹的过程，正义学承认结构而把比较交给分配规则。

还有一句必须写在这里，它是本文与前面所有说法的分水岭：删掉本文这套清点之后，设对不是变得难查，是不存在。因为「另一侧是什么」只有在有人去问的时候才成为一样东西；不问，它就只以结

论的形式留下影子——「他进步了」。设对是被这一次追问造出来的一类对象，不是一个原来就摆在那里、只是没人去看的东西。

\### 三家各自为什么到不了

这一步不是三家疏忽。一个领域看不见什么，通常不是它的缺陷，是它的问题结构的影子——三家各自的核心问题，恰好使这一步成为不必要的。

体育学到不了，是因为它已经把另一侧做成了公共设施。秒表、泳池长度、起跳板的弹性、风速的上限——竞技体育花了一百多年把比较的条件标准化到极致，正是为了让另一侧不必每次重新设置。在一个条件被彻底标准化的领域里，追问「这一次和谁比」显得多余：和所有人比，和所有时候比，这就是纪录这个制度的全部意义。**它到不了这一步，恰恰因为它在这一步上做得太成功了——直到条件本身被改动的那一刻，而那一刻在一百年里只发生过几次。第十节跑的正是那几次之一。**

历史学到不了，是因为它把比较本身列为需要防范的动作。它的纪律是禁令：不要拿今天的范畴去读昨天。一门以「不要随便比」为职业训练的学问，不会去发展一套「比的时候另一侧是什么」的清点方法——因为它的正确回应是不比，而不是比得更清楚。**它到不了这一步，是因为它已经在这一步之前拐弯了。**

正义学到不了，是因为它的问题是分配而不是记账。罗尔斯要回答的是「基本结构该按什么原则安排」，那是一个规范问题；帕菲特发现非同一性之后，整个讨论转向的是「以人为中心的道德是否需要修改」，那仍然是一个规范问题。**没有人问「那么记录上该多一栏什么」，因为记录不是这门学问的对象。**它交出了本文最需要的那件材料，然后把它用在了别处。

三家合起来才够，而且顺序是死的。没有体育学那种把条件标准化到可以被整体更换的领域，就看不见更换这件事。没有历史学的禁令，就不会去问一次比较凭什么成立。没有正义学的非同一性，就想不到另一侧可以是空的。少任何一家，剩下两家给出的都是一个更弱的版本——要么是「测量口径变了」，要么是「不该跨时代比」，要么是「归因很难」。

六 三类设对

本文不用两轴四格。这件事的结构不是两根轴交出来的四格，是一条按「另一侧站着什么」分出来的三分。

甲类·实对。 另一侧是一个可指认的具体对象：同届的考生、去年有记录的自己、上一届的班级、榜上的某个人。它的判据很简单——**能不能报出名字或编号**。

乙类·构对。 另一侧是一个统计构造：同起点学生的期望值、这个年龄段的常模、模型预测的应有水平。它可以被算出来、被复算、被写进报告，**但它不是任何人**。增值评价的整套方法正是建立在这一类上：把一个学生的实际表现与「与他同起点的学生平均会达到的水平」相比，差额记为学校或教师的贡献。

丙类·虚对。 另一侧在替代方案里不存在：那个「若无此事的他」不是他。这一类正是第四节那件材料指出的那一类。

三类的关键差别在于**追问下去会得到什么**：

- 对实对追问，得到一个名字，然后可以问「他和你可比吗」；
- 对构对追问，得到一个模型，然后可以问「这个模型的假设成立吗」；
- 对虚对追问，得到一个停顿——**问题本身没有对象**。

而三类在记录上完全一样：都写着「进步了」。

这三类不是程度，是种类。 一个判断不会「有点实对」；它要么另一侧站着一个人，要么站着一个人造，要么什么都没站。而误判的方向是固定的：**虚对最容易被读成实对**，因为结论的形式完全相同，而没有人去追问另一侧。

七 为什么要打第三类：两条签名与两条误诊阻挡

一个分类若指向的那一类是谁都看得见的坏，这个分类就是多余的。丙类值得单列，是因为它满足两条相当苛刻的条件。

签名一：结论完全可读，而另一侧不可查。

虚对不产生任何异常。「他因为读了这所学校而变好了」是一句语法完整、语义清晰、情感真挚的话，它在家长会上、在校庆演讲里、在校友访谈里天天被说出来，没有一次会被打断。要发现它是虚对，你必须专门去问那一句——而没有任何一份记录格式要求问它。

签名二：更换设对不产生任何登记动作。

这一条是第十节那次回溯检验的直接来源。当比较的基准发生了变化——规则改了、器材换了、题型变了、评分标准调整了——被比较的两侧不再是同一类东西。而记录制度的常规处置**不是给原来那张表加一列写明基准**，而是另起一本账，或者干脆什么也不做。原表照旧、数字照旧、排序照旧，唯一变了的那样东西不在表上。

误诊阻挡甲：不是所有不写出另一侧的判断都是虚对。

有一类判断是真正的绝对判断：「他能连续游完五十米」「他能独立完成这道题」。这类判断不需要另一侧，因为它问的是一个二值事实。**把它们算进虚对是误诊。** 判别很简单：如果这句话在世界上只剩他一个人时仍然有意义，它就不是比较性判断。

误诊阻挡乙：更换设对本身不是坏事，有时正是必须做的。

第十节那个例子里，禁止某类器材是一次正确的更换——它恢复了比较的意义。本文不主张不许换基准；**本文主张的是换了之后要留下一条可以被清点的痕迹。** 把「换基准」当成过错，是对本文的误用，第十八节会把这一条写死。

八 一句不含程度词的问话，与它在五个场合的读数

上面的分析要能用，必须压成一句可以拿去问记录、而不是拿去问人的判断的话。本文给出的是：

这份记录上，与他被比较的那一个，是谁？

这句话里没有「充分」「真正」「实质性」「合理」这类词。它要的是一个名字、一个模型、或者一个「没有」。它的答案不取决于评价者的标准，只取决于记录里有没有那一栏。

把它放在五个互不相干的场合跑一遍，答案必须各不相同。

场合一，升学录取。 答案是一个可指认的人群：同一次考试、同一个招生计划下的其他考生，有名有姓、有编号、可复核。这是最干净的实对，而且它有一处常被忽略的性质——**这个设对每年更换一次，而分数线随之波动**；同一个分数在两年里意味着不同的位置，而档案上只留分数。

场合二，增值评价。 答案是一个统计构造：与他同起点的学生的期望表现。**这一条是查出来的，而且它反过来改了本文。** 本文原以为构对是实对的一个退化形态——先有人群，再从人群算出一个数。查下去发现不是：增值模型里的那个期望值**从一开始就不对应任何一个具体学生**，它是拟合出来的一条线上的一个点，可能落在两个真实学生之间、也可能落在全部真实学生的范围之外。本文因此把构对从「实对的简化」改判为一个**独立的类**——它不是把人群压缩了，它是造了一个原本不存在的对照者。

场合三，世界纪录。 答案是历史上那个具体的人，而**器材规则已经变了**。第十节的回溯检验就跑在这一条上。**这一条也是查出来的，而且它改掉了本文的措辞**，见第十节结果丙。

场合四，个人回顾。 「要是当年没上这个学校，我不会是现在这样。」答案是一个不存在的自己。这是纯粹的虚对，而它是日常语言里最常见的一种教育判断。

场合五，赔偿与矫正。 法律要算「若无侵害本会有状态」，这在实务里叫作基准状态。答案是一个必须被构造出来、且各方会激烈争论的对照。有意思的是：**法律共同体明确知道这个对照是被构造的，并为它设置了举证责任与争辩程序**——而教育评价对同一件事既不设程序也不设字段。

五个答案互不相同，其中第二与第三是查出来的，并且各自削掉或改写了本文原来的一句话。**这正是判据该有的样子：它在不同场合被不同的记录回答，而不是被同一个原则回答。**

九 读数：设对清单，及其清点手册

本文的读数不是一个比率，是一份**清单**。这是有意的：比率会把三类压成一个数，而本文全部的分辨力恰恰在三类的区分上。

读数·设对清单。取一份记录（一学期的评语、一份成长报告、一份评估结论、一张榜单），把其中**全部比较性判断**逐条列出，逐条填这一侧的另一侧是什么，按三类归档；填不出的记为丙类。产出的是一张表，不是一个数。

清单可以附带两个派生量，但它们是描述，不是主读数：**具名率**（甲类条数占比）与**丙类条数**。本文不为它们设阈值——设了阈值，清单立刻会退化成一个可以被优化的指标，而第十八节说明那样做的后果。

清点手册

读数名：设对清单 符号：C

定义：给定一份记录 R 与一段时间窗，列出 R 中的全部比较性判断，

逐条标注其另一侧属甲（可指认对象）、乙（统计构造）、丙（替代方案中不存在）。

产出为一张逐条的表，不折算为单一数值。

什么算一次「比较性判断」：一句包含程度或次序的评价，且去掉另一侧之后该句不成立。

举一个恰好在边界上的例子：「他这学期作业全部按时提交」不是比较性判断（去掉另一侧仍成立）；

「他这学期作业提交情况明显好转」是（去掉另一侧不成立）。

编码规则（六条）：

1. 一句里含两个比较的，拆成两条分别标注。
2. 另一侧只有在记录内部可查时才记甲；需要评价者口头补充的记丙，并在备注里写明补充内容。
3. 统计构造记乙时，必须同时记下该构造的名称与出处（哪一个模型、哪一版常模）；
记不下出处的降为丙。
4. 「与自己比」只有在记录里存在同口径的历史读数时记甲；不存在则记丙。
5. 基准在时间窗内发生过变更的，该条单独标注「基准已变更」，并记下变更日期；
查不到变更日期的，此条不计入任何一类，另列。
6. 清单必须随记录一同保存，不得只保存派生量。

表头：记录来源 | 时间窗 | 判断原文 | 另一侧的内容 | 类别（甲/乙/丙） | 出处 | 基准是否变更 | 备注

不适用：不含任何比较性判断的记录；无法取得原始判断文本、只有汇总结论的记录。

本篇三处引用一致性自查：正文第九节 = 证伪条款第十七节第三条 = 结论第二十节，
三处引用同一份定义、同一套三分与同一份编码规则。

为什么不给一个比率。 一个理由是技术的：三类不在一根轴上，压成比率会丢掉最要紧的那一格。另一个理由是第十八节要展开的：**任何一个关于设对的比率，都可以靠把丙类改写成乙类来做好看——**给虚对配一个模型，它立刻变成构对，数字改善而事情没变。清单挡不住这件事，但清单至少把改写留在纸面上。

十 回溯检验：九十五条世界纪录

证伪条款不能全是纸面的。本文选了一条可以在既有公开记录上直接跑通的：**当比较的基准被更换之后，记录制度实际做了什么。**

为什么选这一例。 竞技游泳有一个别处少见的性质：它的比较基准在一个可以精确指出的时刻被整体更换过一次。二〇〇八年二月起，一类以聚氨酯等非纺织材料制成的连体泳衣被允许下水；二〇〇九年罗马世锦赛前夕，国际泳联投票决定自二〇一〇年一月一日起禁止这类泳衣。也就是说，有一个宽度不到两年的窗口，窗口内外的成绩不是在同一套条件下取得的——而世界纪录表的功能，恰恰是把不同时刻的成绩放在一起比较。

判定规则（在计算统计量之前写死；诚实声明见下）。 本检验不是前瞻性预注册：取数之后我已先看过一次年份直方图，然后才写下判定规则。因此本节按「一条已有数据的回溯检验」报告，证据等级低于前瞻预注册。写死的是判定规则与阈值，不是事后挑出来的结论。

- **R1（孤岛）**：二〇〇八至二〇〇九两年贡献的现行纪录数，须同时大于二〇〇七年及以前全部年份之和、且大于二〇一〇至二〇一一两年之和。
- **R2（长寿）**：在存活满十年的现行纪录里，带器材注记者占比须达到四成。四成这个阈值在计算之前写定；若与年份无关，该占比应约等于带注记者在全体中的占比。
- **R3（注记的地位）**：清点这条注记以什么形态存在——是一个可排序可分组的字段，还是一个角标。

数据。 公开的游泳世界纪录一览，长池与短池、男女及混合接力，逐条含项目、成绩、创造日期，共九十五条现行纪录；其中六条带有非纺织材料泳衣的注记。取数时间为二〇二六年八月十日。任何人可复算。

结果一：R1 成立，而且形状比预期干净。

二〇〇七年及以前的现行纪录：**零条**。二〇〇八至二〇〇九两年：**六条**。二〇一〇至二〇一一两年：**零条**。

那两年是一座孤岛。它前面是空的——十七年以上的老纪录一条不剩；它后面也是空的——禁令之后的头两年一条也没留下来。而夹在中间的那个不到两年的窗口，贡献了今天仍然站着的六条纪录。

结果二：R2 成立，而且量级很大。

存活满十年的现行纪录共十一条，其中六条带注记，占 **54.5%**，远高于四成的阈值，也远高于带注记者在全体中的占比（6.3%）。二乘二表为 $\begin{bmatrix} 6, 5 \\ 0, 84 \end{bmatrix}$ ，Fisher 精确检验双侧 $p = 5.3 \times 10^{-7}$ 。

存活年数的对比更直接：**带注记的六条，存活中位数十七年（范围十七至十八年）；其余八十九条，存活中位数两年（四分位一至四年，最长十四年）。相差八倍以上。**

结果三（对本文不利，且它直接改掉了本文的措辞）：那一栏并非无人记。

R3 的清点结果是：**那六条纪录确实带着一个注记**，注记的含义是「以非纺织材料泳衣创造」，而且带注记的恰好就是那六条、只有那六条——注记与年份完全重合。

这推翻了本文原来的一句话。本文原本要写：「基准变更不留任何痕迹。」这句话是错的，本文已改。修订后的说法是两句：

其一，痕迹的形态不是字段，是角标。 那个注记出现在成绩单元格里，是一个附在数字旁边的标记；这张表没有一列叫作「适用的器材规则」。角标与字段的差别是可操作的：字段可以排序、可以筛选、可以分组统计、可以被下游系统读取；角标只能被人眼看见。**本文的清单要的是字段，而实际存在的是角标。**

其二，制度对基准变更的处置是另立一本账，而不是给原账加一列。 由于禁令后的成绩长期无法超越窗口内的纪录，一份平行的「纺织材料最佳成绩」榜被维持起来，与官方纪录表并行。这个处置很务实，而它的代价正是本文要指出的：**同一件事从此有两套互不通约的记录，而「这一条到底和谁比」这个问题仍然没有一字段来回答——它被回答成了「看你查哪一张表」。**

结果四（也对本文不利）：最强的那句竞争解释，本检验完全无法排除。

那句话是：**其实不就是那两年的泳衣让人游得更快吗？** 这句话是对的，本文不否认它。带注记的纪录之所以长寿，最朴素的解释就是它们被创造时有额外的助力。

本检验排除不了它，也不打算排除它。因为本文要问的不是「那些成绩为什么快」，而是「**当比较的基准被换掉之后，记录制度如何登记这件事**」。这两个问题的答案互不影响：即便泳衣的效应可以被精确量化，那张表上仍然没有一列写着基准。**本文因此把落点从「为什么」整个挪到「怎么记」——**这一挪使本文放弃了对成绩差异的任何解释权，是本文在这一节付出的最大代价。

结果五（方法学上的自我限制）：R1 与 R2 不独立。

孤岛与长寿其实是同一件事的两种说法——正因为那两年的纪录长寿，它们才在今天的现行纪录表上形成孤岛；正因为它们形成孤岛，长寿才被看见。**两条规则合起来只算一条证据，不算两条。** 本文不把它们当作互相印证的两个发现。

一处必须写明的未做。 第十七节条款七要求把这个形状拿到别的项目上去看——田径的跑道材质与鞋底规格、场地自行车的车架规则、跳水的板性规格，都发生过可以指出时刻的整体更换。**本文一个也没有跑。** 只跑一个项目而声称一般规律，是本文在这一节最薄的一处；结果一的那座孤岛目前只是一个项目的形状，不是一条已被复现的规律。写在这里而不是留给读者去发现。

代理变量声明，四条。 ① 现行纪录表是一份幸存者名单，不是全部纪录的历史；本检验能说的只是「哪一批更难被打破」。② 「非纺织材料」这一注记由编纂者添加，不是记录机构的原始字段——这正是 R3 要清点的东西，也是结果三的来源。③ 短池与长池、个人与接力混在一起计数，分项样本过小，本检验不分层。④ 本检验只用于说明记录制度对基准变更的登记方式，不构成对任何运动员成绩的评价。

十一 在别的领域里能据此预测出什么

一个判断若只在提出它的那个场合成立，它就是一句修辞。下面十二处不是类比，每一处都能据此预测出一件具体的、可以去查的事。

一，**分班与分轨**。 预测：一次分班之后，同一个学生的评语里丙类判断的比例会上升——因为他此后被与新班级比，而旧班级的记录不再进入他的对照。可查：取分班前后各一学期的评语，按第九节手册各清点一次。

二，**转学**。 预测：转学生的记录里，「与自己比」这一类会显著多于本校生——不是因为他更需要纵向比较，而是因为他与新同学之间缺少同口径的历史读数，横向比较无从设置。可查：同一所学校里转入生与非转入生的清单构成。

三，**评语与操行评定**。 预测：越是正式、越是要归档的评语，比较性判断的密度越高而具名率越低——正式文书要求「客观」，而写出具体对照者被认为不客观。可查：同一位教师写给家长的私信与写进档案的评语，两份清单并排。

四，**教师评价与增值模型**。 预测：以增值模型为主的评价体系，其全部判断落在乙类；而当模型版本更换时，同一位教师的评价可以在不改变任何教学行为的情况下反向。可查：模型更换年份前后的评价排名相关系数。

五，**学业竞赛与奖项**。 预测：奖项的含金量随参赛人群的构成而变，而奖状上不写参赛人群。可查：同一奖项在不同年份的获奖人数与报名人数之比，与奖状文本的对照——后者不含前者的任何信息。

六，**职业资格与执照**。 预测：以固定分数线发放的执照，其分数线在历史上是从某一届人群的分位数定出来的；一旦定下，人群变了而线不动。可查：任一执照考试的定标文件，看切分点的来源。

七，**医学的参考区间**。 预测：一份「各项指标正常」的化验单，其正常的含义是「落在某个参考人群的中间百分之九十五」，而这个参考人群不写在化验单上。可查：任一化验单的表头，看有没有一栏写明参考区间的来源人群与年份。

八，**竞技体育的纪录制度**。 第十节已跑。预测：凡是发生过器材、场地或规则整体更换的项目，其现行纪录表上都会出现一个年份孤岛，而表上没有一列写明基准。可查：田径的跑道材质与鞋底厚度规定、场地自行车的车架规则、跳水的板性规格。

九，**人工智能的能力评测**。 预测：一个模型「超过了人类平均水平」这句话里的人类平均，来自一批具体的标注者，而榜单上不写这批标注者是谁、怎么招募的。可查：任一主流评测榜的说明文档，看有没有对照人群的构成。

十，**绩效考核与末位淘汰**。 预测：强制分布的考核制度把每个人的设对锁定为同一考核单元的其他人，因此改变单元边界就能改变全部人的评价而不改变任何人的工作。可查：一次部门重组前后的考核分布。

十一，历史评价与追授。 预测：对一位历史人物的重新评价，其争论的实质常常是设对的更换（与同时代人比，还是与今天的标准比），而争论双方通常不把这一点写出来。可查：任一次重评的公开论辩，看双方有没有明说各自的对照。

十二，法律的赔偿基准。 预测：赔偿争议中双方投入最多的部分，是构造与攻防那个基准状态，而这一部分在判决书里往往被压缩成一句结论。可查：任一份涉及可得利益损失的判决书，看基准状态的论证长度与结论长度之比。

十二 三处反过来削弱本文的约束

跨领域不是只往外送。有三处是别的领域反过来给本文加的约束，每一处都真的削掉了本文原本打算写下的话。

第一处，来自体育学：绝对判断是存在的，而且是必需的。

一个人能不能连续游完五十米，这件事不需要另一侧。运动技能的教学里有大量这样的判断——能不能翻身、能不能独立完成一整套动作、能不能在水下睁眼。它们是二值的、是可教的、是学习者自己也能确认的。

本文原本会写下这样一句更强的话：「**一切教育判断都在另一侧放了一样东西。**」这句话现在必须删掉。适用范围随之缩小：**本文只处理比较性判断**，而第七节的误诊阻挡甲正是为此而设。

这一处削弱的代价不小。它意味着本文不能主张「教育的本质是比较」——本文能主张的只是：**在教育中那些被用来分配机会的判断，绝大多数是比较性的**；而这一句需要经验支持，本文没有提供。

第二处，来自历史学：本文自己的处方也踩在时代错置的禁令上。

本文要求把设对写下来。但写下来的那个设对，是**今天的人用今天的范畴写下的**。一份二十年前的成绩单，即便当年就写明了「与本届三百二十名考生比」，今天的读者仍然要问：那三百二十人是怎么被选进来的、那次考试意味着什么、那个「比」在当时是什么含义。**登记设对不能保证跨时代可比**。

本文原本会写下「登记设对可以使不同时期的评价重新可比」。这句话必须削弱为：**登记设对使得「不可比」这件事变得可查，而不使它变得可比**。这是一个弱得多的主张，但它是本文能守住的那一句。

第三处，来自正义学：有些设对不该被写出来。

这一处动的是价值排序，也是本文让步最大的一处。罗尔斯的基本结构说提醒的是：把一个学生与另一个具体学生并列写进他的档案，本身就是一次伤害——它把一个人变成另一个人的度量衡，而且这份文书会跟着他走很多年。第十八节的伦理条款正是从这里来的。

本文原本会写下「每一次比较性判断都应当在记录中写明其另一侧」。这句话必须改成：**每一次比较性判断都应当在记录中写明其另一侧的类别与来源，而当另一侧是具体个人时，只登记类别与人群构成，不登记那个人**。处方因此从「写出来」退到「写出它是什么类型的东西」——本文的可操作性下降了一档，而这一档是被正义学收走的。

十三 与最近的几种说法的分界

下面十条说法都可能被读者与本文混为一谈。逐条写清它说到哪一步、差别落在哪、以及一个能把两者分开的判定形式。

一、**潜在结果框架**（Rubin 1974; Holland 1986）。它说到哪一步：一个单位在受处理与不受处理两种情形下各有一个潜在结果，因果效应是两者之差；由于只能观察到其中之一，因果推断是一个缺失数据问题。**这是本文在方法学上最危险的一位**，因为它把「另一侧」形式化到了极致——那正是本文的丙类要处理的东西。

分离线在于**那个单位是不是同一个**。潜在结果框架预设同一个单位 i 在两种处理下都存在， $Y_i(1)$ 与 $Y_i(0)$ 说的是同一个 i ；非同一性说的是，在一类足够深的介入下，另一侧那个 i 不存在——不是它的结果缺失，是它没有下标。判定形式：**若一切设对都能被写成同一单位的另一个潜在结果，则本文可整个并入这个框架，只是换了个说法。若存在一类介入，其替代方案里的受益者不是同一个单位——同一性本身被介入改变——则潜在结果在那一类上没有定义，而本文的丙类正是为它保留的位置。**帕菲特那个例子恰好就是这样一类。

二、**参照群体理论**（Merton & Kitt 1950; Hyman 1942）。它说到哪一步：人评价自己的处境时，参照的不是绝对水平，而是他所选取的那个群体；参照群体变了，同样的处境会带来相反的满意度。**这是本文在经典层最近的一位。**

分离线在于**谁在设**。参照群体理论研究的是人自己选择跟谁比，那是一个心理与社会过程，可以问当事人；本文研究的是记录制度设置了跟谁比，那是一个记账动作，只能问文书。两者可以完全脱钩，而脱钩本身可查。判定形式：**取一批学生，分别问他们「你觉得你在跟谁比」，再按第九节手册清点他们档案里的设对。若两者高度一致，本文只是参照群体理论的文书版；若系统性不一致——尤其是当事人报出的是具体同学而档案里全是统计构造——则两者管的是两件事，而后者才决定机会分配。**

三、**常模参照与标准参照**（Glaser 1963）。它说到哪一步：测验分为两类，一类把个体放到人群分布里定位，一类对照一条事先规定的能力标准。这是教育测量的教科书二分，用了六十多年。**它是本文在应用实务层最近的一位。**

分离线有两处。其一，这个二分只有两格，**装不下丙类**——它默认另一侧要么是人群、要么是标准，没有为「另一侧不存在」留位置。其二，也是更要紧的一处：**它把标准参照当成没有比较对象的判断，而标准本身通常是从某个人群的分位数定出来的。**判定形式：**取任一自称标准参照的考试，查它的定标文件。若其切分点的来源是一个人群的分位数，则它是一次伪装成标准参照的常模参照，本文的三分能把它读出来而那个二分读不出来。若切分点来自一份与人群无关的任务分析，则那确实是标准参照，本文在这一类上不适用。**

四、**社会比较理论**（Festinger 1954）。它说到哪一步：人有评价自己的驱力，在缺少客观标准时会与他人比较，且倾向于选择与自己相近的人。分离线与第二条同类而更靠前：它是关于**驱力**的理论，本文是关于**文书**的清点。判定形式：在一个完全不允许当事人知道他人成绩的制度里，社会比较的效应应当消失，而本文的清单不变——因为档案照旧在设置设对。

五、**增值模型**（Sanders & Horn 1994; Braun 2005 的批评）。它说到哪一步：把学生的实际表现与模型预测的期望表现相比，差额归给学校或教师。它是构对最成熟的实现，本文欠它一笔。分离线在于**它把那个期望值当作方法，本文把它当作对象**：增值模型的全部争论围绕「这个期望估得准不准」，而本文问的是「这个期望是不是一个人」——两个问题的答案互不影响。判定形式：一个估计得极准的增值模型，在本文的清单上仍然整列记乙；若有人主张估准了就等于设对成立，请他回答那个期望值对应哪一个学生。

六、**时代错置的禁令**（Skinner 1969; Butterfield 1931 的辉格史批评）。它说到哪一步：用后来的范畴去读前人，会得到一整套看似合理而实为虚构的东西。分离线在于**方向**：时代错置的禁令是**禁止**一种比较；本文不禁止任何比较，只要求把比较的另一侧登记下来。判定形式：一次明知不可比而仍然作出的比较，按时代错置的禁令应当被判为错误，按本文只需在清单上标注「基准已变更」并保留——**两者对同一份材料给出相反的处置**：前者要求删掉，后者要求留下并标注。

七、**非同一性问题**（Parfit 1984）。它是本文的推翻材料，不是竞争者，但必须写清界线。它在道德哲学里回答的是「我们对尚未存在的人有没有义务」，落点在义务；本文取它的形式后果——**一部分比较缺少另一侧**——落点在记账。判定形式：若一切丙类都能被归结为道德义务问题，本文多余；若存在一类丙类判断与义务无关（例如一次纯粹的能力评估），则本文说的是另一件事。

八、**排名的自我实现**（Espeland & Sauder 2007, 2016）。它说到哪一步：一个排名一旦发布，被排名者会为了排名而改变行为，于是排名逐渐制造出它所宣称测量的那个东西。分离线在于**回路的有无**：排名的自我实现需要被排名者知道排名并作出反应；本文说的设对**不需要任何人知道**——一个从不公布的内部评价，其设对照样在设置，照样决定资源配置。判定形式：在一个评价结果完全不向被评价者披露的制度里，自我实现的效应应当消失，而本文的清单构成不变。

九、**制造人的种类**（Hacking 1986, 1995）。它说到哪一步：一个新的分类被造出来之后，被分类的人 would 按这个分类改变自己，分类与人互相塑造。分离线与第八条同源：回路效应要求被分类者知道；而且更要紧的是，**哈金处理的是「他属于哪一类」，本文处理的是「他被他比」**——一个人可以在不改变任何分类归属的情况下更换设对（换一个班就够了）。判定形式：一次不改变任何分类标签、只改变分组边界的调整，若引起评价结果的系统性变化，则本文对回路效应不足以解释。

十、**剥掉全部名词之后的形式**。把本文的命题去掉教育与体育的一切名词，剩下的是：**一次判断必须在另一侧放一样东西，而这一样东西的身份从不进入该判断的记录；当它被更换时，记录不变而结论的含义变了**。这个纯形式的结构在上游有名字。计量学里叫**溯源与参考标准的变更**——一次量值的意义依赖于它所溯源的那个标准，标准换代时旧值需要换算。会计学里叫**会计政策变更与追溯调整**，准则要求披露变更的性质、原因与影响数。统计里叫**基期变更与指数链接**。

三者当中，会计学那一条对本文最不利，必须正面说：会计准则已经要求在报表附注中披露政策变更及其影响数——也就是说，「基准变了要登记」这条处方在另一个领域里已经是强制规范，本文不是第一个想到它的人。本文能主张的只剩两件：其一，教育与竞技的记录制度**没有**这条规范；其二，这三个上游领域处理的都是**同一个对象在不同基准下的换算**，而本文的丙类说的是**另一侧根本不存在**，那是换算处理不了的一类。判定形式：若一切设对变更都能通过换算或链接处理，本文可并入基期变更那一族；若存在一类变更，其变更前不存在任何可用于换算的重叠观测（因为另一侧的那个人不存在），则本文的丙类在那一族之外。

与同题另外六篇的分界。 本文与另外六篇同题稿相邻到必须逐篇分开，而逐篇的对照表在第十四节；此处只写那条贯穿全部六篇的共同分离线——**那六篇的落点全部在被评价的那一侧**：某样东西在这把尺子上读不出来、算不到、或者根本没被放进字段。**本文的落点在比较的另一侧**：那把尺子的另一端放着什么。两者正交，而且叠加时有具体形状——一族在口径下分不开的人，可以有完全不同的设对（一个被与同班比、一个被与常模比），而设对的差别不进入任何读数。判定形式：**在同一份记录上分别按七篇的手册标注。若「那六篇各自读不出的那一类」与「本文另一侧记丙的那一类」两组标注的重合度低于百分之四十，则七篇处理的是七件事；若高于百分之七十，则本文应当并入其中最接近的那一篇。这次清点本文没有做，是本文欠下的一笔。**

与本文最接近的是第一条（潜在结果框架）。 它不能吸收本文的理由只有一条：它的整套形式要求另一侧那个单位存在，而本文的丙类正是为它不存在的那一类保留的位置。

十一、参照点与锚定（卡尼曼与特沃斯基，一九七九；特沃斯基与卡尼曼，一九七四）。这是全部十一位里最近的一位，本稿修订前漏掉了它，而它就长在本文的题目上。 它说到哪一步：人对结果的评价不是绝对的，而是相对于一个**参照点**判为得或失；而**参照点不由情境给出**，它可以被措辞、被先前的状态、被一个任意抛出的数字（锚）挪动。四十年的实验证据说的正是「另一侧放了什么，决定了这一侧被读成什么」。

分界只有一条，但它与本文对潜在结果框架那一刀完全同形，因此可以一并判：**参照点永远存在。** 它是同一把标尺上的一个值——可以被挪、被操纵、被误设，但它总在那里，因而总能被写下来、被比较、被纠正。本文的丙类说的不是参照点被设错了，是**另一侧根本没有那个下标**：不是「与一个不合适的对照相比」，是「这句话没有对照可言而仍被说成了比较句」。前者的处方是把参照点校准，后者的处方是登记「这里没有对照」——两者在记录上的形状不同，一个是改一个值，一个是加一条缺项。

还有一层顺带的分别：参照点研究关心的是**评价者的心理**（他把哪里当作原点），本文关心的是**记录的形态**（那份文书上另一侧写着谁）。同一次比较，参照点可以在评价者心里而记录上一片空白，也可以记录上写满而评价者根本没看。两者可同批测且互不蕴含。

判定形式：取一批被判为丙类的比较性判断，向评价者追问「你是跟什么比的」。若**他们总能报出一个具体的值或状态**，则**参照点一路对**，本文的丙类只是**参照点未被记录**；若有一类判断，评价者自己也报不出另一侧是什么、而判断照样作出并照样被采纳，则**本文对**。这一条便宜，可在任何一次评语式评价的复盘访谈里做。

十四 同题七篇的互切

本文与另外六篇同题文章出自同一批工作：七篇都问「教育是什么」，七篇都以「某个口径读不到、算不到或根本没放东西的那一类」作答。本文是七篇里唯一把落点挪到比较另一侧的一篇，正因如此更需要把七篇摆在一张表上逐列分开。

篇	三门学科	它命名的那一类	读数	相对于什么被定义	对象是什么	有没有时间结构
它读不出来的，它也造不出来	工程学·生物学·艺术学	未张成能力	张成重合率 λ	入判口径的边界	能力空间里的方向	无
那道题还没有被出出来	符号学·逻辑学·数学	界题	分法未定率 u	已出题目的集合	尚未被出出来的题	无
同许	混沌学·舞蹈学·地理学	同许族	带宽·同许率·分离视界	可用观测的粒度	人与人之间的配对	有，会自己散开
不计的那一遍	戏剧学·经济学·法律学	不计遍	零计比 z ·首计序 k	计数装置的字段表	做过的那些遍	有，被记录即消灭
本来就是这样	圣经学·佛学·道家	未判之域	判形滞后期	传承体系的留造分配	所传之物的一部分	有，被指出即消失
先空着的那一处	营养学·水力学·园林学	无认空	留空清单·认领时刻	责任清单的认领状态	接受端的位置	有，被认领即开始填
设对	历史学·正义学·体育学	虚对	设对清单	比较判断的另一侧	被比较的那一个	无，它根本不存在

七篇的对象类型互不相同——方向、题、配对、遍、内容的一部分、位置、被比较的那一个——这是最干净的一刀。七类可以同时非空且互不重叠：一门课的成果从不进入任何判定（未张成非空），而它的考卷每道题只考一项技能因而能分开一切做法（界题为空）；同一批人分数彼此差得很开（同许族为空），而他们为达到这个水平做过的重复里九成不产生条目（不计遍非空）；这门课的教材里有一大段从没有人问过“必须原样还是可以重讲”（未判之域非空），而课上每个空位都写在某位教师的清单上（无认空为空）；而给这个班打的每一个分数，另一侧站着的是一个真实存在的对照班（虚对为空）。七个读数在同一批人身上可以取到任意组合。

判决性对照：一个动作，七种不同的预测。 取一批本来不进入分配决策的记录，只把它接进判定，其余一概不动：

- **未张成能力**：被切出的那一类整批更换成员，总量不变。
- **界题**：分法未定率不变——接进判定的记录并不增加能把做法分开的题，除非它恰好只考一项技能。
- **同许族**：带宽变窄、同许族变小，而分离视界不随之改变。

- **不计遍**：那批重复的构成功能下降——记了就换格。
- **未判之域：触发一次判形**——被接进判定的那部分内容从此有了留造归属，未判之域缩小一块，而其余部分不动。
- **无认空**：外部来源的进入率下降——一旦有了责任人，填它的通常就是认领它的人。
- **虚对：读数不变**——设对清单问的是另一侧站着谁，接进判定的是这一侧的记录，两者不相干；除非新记录本身规定了对照来源。

七条不能被同一次观测同时满足：第二条与第七条说读数不动，第三条说读数变小，第五条说读数缩小一块，第四条与第六条说被记的东西性质改变，第一条说总量守恒而成员更换。**任何一次真实的口径变更都会至少判掉其中两条。**七篇因此是可以互相证伪的七个判断，不是一个判断的七种说法。

还要写明读者最该起疑的地方，本文不绕过：**这七篇出自同一位作者、同一批工作、同一副骨架**（三家现成读法 → 三家共有前提 → 一件来自其中一家自己的推翻材料 → 命名一类 → 一套辨别装置 → 一句不含程度词的判据 → 一个可清点的读数 → 一次预注册或回溯检验 → 逐条分界 → 编号证伪）。骨架相同不构成七篇内容相同的证据，但它确实是七篇最可能出错的地方：**一副好用的骨架会让七个不同的问题看起来长得一样。**判据交给读者：上表七列里，若有两篇能在五列以上取到相同的值，那两篇应当合并。现在最接近的一对是**不计遍与无认空**（同为“被登记/被认领即消灭”的形状，且两篇的经验材料取自同一个公开平台），它们在七列里相同的有两列；其余任意两篇相同不超过两列。这条判据可以拿去核对，也可以拿去推翻。

十五 所有人共同站着的那块地

把上面十种说法逐个问一句「它把什么当作给定，因此看不见什么」，答案有一个共同点。

潜在结果框架把「单位在两种处理下都存在」当作给定，因此看不见同一性本身可以被处理改变；一旦承认它可以被改变，因果效应这个定义在那一类上就没有对象。

参照群体理论把「比较是当事人做的」当作给定，因此看不见文书替他做了一次；一旦承认文书也在设，那么改变一个人的处境就不必改变他的任何想法。

常模参照与标准参照把「另一侧要么是人群要么是标准」当作给定，因此看不见第三种；一旦承认第三种，这个二分就不再穷尽。

增值模型把「期望值是一个可估的量」当作给定，因此看不见它不是一个人；一旦承认它不是人，「教师的贡献」这个说法就失去了它所声称的那个对照。

时代错置的禁令把「不可比就不该比」当作给定，因此看不见不可比的比较也可以被登记而不必删除。

排名的自我实现与制造人的种类都把「要有回路就得让人知道」当作给定，因此看不见不公开的评价照样在分配。

会计学的政策变更披露把「变更前后是同一个对象」当作给定，因此看不见有一类变更没有可用于换算的重叠。

这七行背后是同一个被当作给定的东西：

凡是一次比较，另一侧总有一样东西站着。

说得更具体一点：**所有人都在争标准该定在哪儿、该跟谁比、比得公不公平，而没有人问过——这一次比较的另一侧，是不是真的有东西。**互相批判的各派在这一点上从不交锋：主张绝对标准的与主张相对标准的，主张个人努力的与主张结构补偿的，全都默认另一侧是给定的，分歧只在于该给定什么。

这条地基不是某一派的立场，它是各派共同站着的地面。而本文提出的那个东西，正是从这块地的缝里掉下去的那一类：**在丙类判断里，另一侧空着，而结论照常作出、照常归档、照常用于分配。**它不是记录得少——分数、日期、签名，每一栏都记得很认真；它是「另一侧是什么」这件事压根不算一样东西，所以没有任何一栏为它而设。

「没有人辨识过这件事」的真正来源不是没有人看，是**看的人都站在它上面。**

十六 不适用的边界

本文明确不适用的六种情形：

一，**非比较性判断**。第十二节第一处已把这一处让给了体育学：二值的、绝对的、当事人自己也能确认的判断，不设另一侧，也不该被算进清单。

二，**只有汇总结论的记录**。清单要求逐条取原始判断文本；只保留了「优秀」「合格」这类结论而没有判断原文的记录，算不出清单。

三，**设对被有意隐去的场合**。匿名评审、盲评、双盲试验——这些制度故意不让评价者知道另一侧是谁，而这是为了公正。本文的清单在这类场合会大量记丙，但那不是缺陷，是设计。清单必须为这类场合单设一栏标注，否则会把好制度读成坏制度。

四，**跨时代的可比性**。第十二节第二处已把这一处让给了历史学：登记设对使不可比变得可查，不使它变得可比。凡涉及跨时代评价的推论，本文一概不作。

五，**本文不判断一次设对是好是坏**。一个被有意设置的构对（用统计构造代替具体同学，正是为了避免伤害）与一个因疏忽而落空的丙类，在清单上都记非甲。要分开它们，需要问「这个设对是谁定的、为什么这么定」——那是一个关于制度意图的问题，本文没有提供回答它的工具。**这是本文交出去的第一个洞。**

六，**本文说不出丙类有多常见**。第十节跑的是竞技体育的记录制度，不是教育记录。教育里丙类的占比是多少，本文一次也没有清点过——第十七节条款三正是为此而设，而它至今是设计不是发现。**这是本文交出去的第二个洞，也是最大的一个。**

第三个洞：本文的判据依赖「比较性判断」这个界定，而一句话算不算比较性判断，在边界上要靠人判断。第九节的编码规则给了一个可操作的检验（去掉另一侧是否还成立），但它挡不住所有争议。本文能做的只有把判断原文强制留在表头，让第三方可以复核。

十七 证伪条件

先写出最难反驳的那句竞争解释，再让证伪条件排除它。

最强竞争解释：其实不就是记录写得简略吗？ 也就是说，没写出另一侧只是因为文书要节省笔墨、为大家心里都清楚、因为写出来太啰嗦。按这个解释，只要要求评价者补写，问题就解决了；本文所谓的「三类」只是补写完整程度的三个档次。

排除它的方式：**若只是简略，那么补写之后三类应当全部收敛为甲类或乙类。** 本文预测不会——**丙类补不出来**，因为它的另一侧不存在，补写只会把它改写成个构对（临时造一个对照），而这个改写动作本身可以被看见。

条款如下，八条。

一（主条款，机制证伪）。 取一批已归档的评语，要求原评价者对其中的每一条比较性判断补写另一侧。控制补写时间与工作量之后，若补写后丙类归零、且补写内容在第三方复核下均可追溯到记录内已有的对象，则本判断为伪——届时「没有另一侧」应归因于文书简略。样本要求：同一学校体系内 $N \geq 6$ 位可比教师，不得以两所学校的对比充数。

二（正向读数，顺序预测）。 一次分组边界的调整（分班、分校区、改考核单元），**设对的更换应当早于任何读数的变化**。若在纵向记录中设对更换的中位提前时间的置信区间包含零或为负，则本判断为伪。用顺序而不用相关，因为相关太容易被巧合满足。

三（低成本、可当场执行的一条）。 这一条本文认为全篇最该先做：取任意一所学校一个学期的全部学生评语，按第九节手册清点一次设对清单，报告三类构成与具名率。**全部材料已经存在，不需要新的采集，不需要前瞻设计，一周之内可完成。** 若清单中比较性判断的占比低于两成，则本文所处理的现象在教育文书里不是主要形态，本文的适用面须大幅收窄。

四（与潜在结果框架的分离）。 若一切被记为丙的判断，都能被重述为同一单位的另一个潜在结果，则本文的第三类不成立，本文可并入因果推断的既有框架。

五（与常模参照二分的分离）。 取十份自称标准参照的考试的定标文件。若其中切分点来自人群分位数的不足两份，则「标准参照多为伪装的常模参照」这一预测为伪，本文第十三节第三条的判定形式失效。

六（与排名自我实现的分离）。 在一个评价结果完全不向被评价者披露的制度里，若设对更换不引起资源分配的系统性变化，则本文的机制需要回路，与自我实现不可区分。

七（回溯检验的推广）。 第十节的形状应当在其他发生过基准整体更换的项目上重现。若在三个以上此类项目中，现行纪录的年份分布不出现与更换窗口重合的孤岛，则第十节的结果是这一个项目的特例，不支持一般主张。

八（回溯检验复算）。 第十节的数据来源、判定规则、编码与统计脚本已全部写明，任何人可复算。若第三方复算所得的孤岛构成与本文报告不一致，则本文的编码有未写明的自由度。

若本判断被证伪，我们会学到什么。若条款一失败，说明另一侧其实一直在评价者心里、只是没写下来，那么改进的方向就回到「要求写完整」这条便宜得多的老路上——这对实务是好消息。若条款三失败，说明教育文书里比较性判断本就不多，那么整个问题的重要性被高估了，而本文应当降级为一篇关于竞技记录制度的文章。两种失败都会让事情变简单，这一点本文必须说在前面。

十八 伦理与证据边界

设对清单是一份可以被拿去考核人的东西。它一旦进入教师评价、学校评估或资源分配，会立刻产生真实的不利后果。因此写死三条。

一，**这份清单指的是文书，不是人。** 它描述的是一份记录里比较性判断的构成，不得作为任何评价者的能力、态度或责任心的评级使用。**不存在「丙类率高的教师」这种说法**——丙类多，多半说明这位教师在写档案时避免了把学生与具体同学并列，而那正是第十二节第三处所要求的。

二，**不得为把清单做好看而制造具名。** 这一条是本文全部伦理件里最要紧的。具名率上升有两条来路：一条是评价确实变得可追溯，一条是有人把学生与具体同学的对照写进了档案。**后者在清单上与前者完全一样，而它是一次伤害。** 因此明确禁止：不得为提高具名率而在学生档案中登记具体的对照个人；正确的做法是登记设对的**类别与来源**（哪一个人群、哪一版模型、哪一年的常模），不登记那个人。

三，**已经按这份清单做出不利安排的，必须公开编码规则并给出复核通道。** 任何据以评价教师、调整制度或分配资源的决定，都必须同时公布所用的记录范围、时间窗、编码规则与判断原文，并允许受影响者按同一手册复算。

反噬预言。 这份清单一旦进考核，最省事的达标方式恰恰会摧毁它要保护的那个东西：**给每一个丙类判断临时配一个模型。** 它立刻变成乙类，具名率与结构构成同时改善，报表立刻好看——而那个模型可以是任意的、事后造的、从不被检验的。第九节之所以拒绝把清单折算成一个比率，正是因为比率会让这条路径变得几乎免费。清单挡不住它，只能让改写留在纸面上。**我看不到出口。** 本文能做的只有把这一条写在这里，让任何采用者先读到它。

十九 自反：本文的设对是谁

本文提出了一个判据，那就得允许这个判据回头指向本文自己。

本文通篇在作比较性判断：本文说它比现有的读法**多**指出了一件事，说第十三节那十位邻居**都差最后一步**，说这个问题**没有人**辨识过。按本文自己的定义，这些全部是比较性判断，因此本文自己也有设对。那么本文的设对是谁？

答案是另外六篇同题的文章。它们与本文出自同一批工作、问的是同一道题，逐篇的分别已列在第十四节。本文说「本文的落点在另一侧」，这句话的另一侧站着的正是那六篇。

按第九节的手册，这个设对记**甲类**——它是可指认的，有篇名、有日期、有构念名。到这里为止，本文是干净的。

但接下来的一步不干净。那六篇的存在，本身是被同一条写作线设置出来的：同一道题被连着出了七次，每一次换一组学科。于是本文的设对不是独立于本文而存在的对照，是**与本文由同一只手在同一天造出来的**。用本文第十节的话说：这不是一次实对，是一次在**同一套条件下自制的对照**——它在形式上可指认，在功能上不构成检验。

这个定位有三个具体后果，本文照单接受。

后果一。本文声称的「多出的那一步」，其可信度不高于「同一条线上的另外六篇没有走那一步」这个事实，而这个事实是本文自己安排的。要让这一步真正成立，需要的对照是**别人在不知道本文的情况下处理同一个问题**——而那样的对照，本文一个也没有找到并写进第十三节。第十三节的十位邻居是真实的、独立的，但它们没有一位是在回答「教育是什么」。

后果二。本文批评评价制度不登记设对，而本文自己直到这一节才把设对写出来。前面十七节里的每一次「更」「多」「差最后一步」，都是在没有登记另一侧的情况下作出的比较性判断。**按本文自己的清单，本文正文的具名率在第十九节之前为零。**

后果三，也是唯一可执行的补救。本文应当把第十三节末尾那次未做的清点做掉：在同一份记录上按七篇的手册各标注一次，报告重合度。第十三节已经写明这次清点没有做，并把它列为欠账；这一节把它升格为条件——**若本文发表之后一年内这次清点仍未做，那么本文与那六篇的分界就只是几段说明文字，读者有权据此把七篇当作一篇的七次改写。**

还有一层更不客气的自反。本文第十八节禁止「为把清单做好看而制造具名」，而本节所做的事，形式上正是那个动作：本文为自己的比较性判断补上了一个可指认的另一侧。**唯一的差别是本文同时写出了这个另一侧是自制的。**这个差别是不是足够，本文不能自己裁定。

二十 结论，与一条写死日期的赌注

本文只主张一件事：

一切「他变好了」都在另一侧放了一样东西；那一样东西有三种，而三种在记录上长得一模一样；其中一种在替代方案里根本不存在，而这一切都不进任何一栏。

它可以被分三层引用，各层的可靠性递增。

第一层，设对这个构念本身——比较的另一侧是一类可以被单独清点的对象，它有三种，且第三种在结构上不可补写。这一层最强也最易倒：第十七节条款一与条款四直接指向它，而第十节的回溯检验没有在教育记录上跑过。

第二层，三分作为分析工具——即便设对的本体地位可疑，「另一侧是可指认的人／是统计构造／不存在」这三分仍可用来分类现有的评价文书，尤其是把乙类与丙类分开。这两类在结论的形式上完全相同，而它们要求的处置完全相反：乙类要问模型的假设，丙类要问这句话有没有对象。

第三层，一条不依赖前两层的经验主张：在发生过基准整体更换的竞技项目里，记录制度对更换的处置不是给原表加一列，而是另立一本平行的账；而在教育记录里，连这本平行账也没有。前半句本文跑过了——九十五条现行纪录、一座两年宽的孤岛、一个角标而非一个字段；后半句最容易检验，随便取一份学籍表数一数字段就知道，也最稳，而且它独立于本文其余全部主张成立。

一条写死日期的赌注。

到二〇三三年十二月三十一日之前，将有至少一份公开的教育评价规范或学籍数据标准，在其正式字段表中设置一个专门记录「本次评价所依据的比较基准」的字段。该字段须要求填写基准的类别与来源，而不只是一个自由文本备注。

什么不算命中，写死四条：

- 1\ 只在研究方案或试点文件里出现而未进入常规记录格式的，不算。
- 2\ 只设一个自由文本的「备注」栏而不规定填写类别与来源的，不算——那是角标不是字段，第十节结果三已把这两者分开。
- 3\ 只要求机构自行说明「本次评价参照了相关标准」的，不算——这类表述不可核。
- 4\ 由本文作者或其合作者参与制定的，不算。

最后回到开头那句话。

本文不能告诉你那个孩子到底有没有进步。本文能说的是：说他进步了的那个人，在另一侧放了一样东西；那样东西可能是他的同学、可能是一条从别人身上算出来的线、也可能是一个并不存在的、若无此事的他。三种说法在成绩单上写出来是同一句话。

而那份成绩单之所以看上去客观，恰恰是因为它只写了一侧。

注释

〔一〕本文使用的「比较性判断」一词，指一句包含程度或次序的评价，且去掉另一侧之后该句不成立。第九节的编码规则给出了一个可操作的检验，并举了一个恰在边界上的例子。

〔二〕「设对」中的「设」字取设置义，不取假设义。本文强调的是它被放置这一动作，而不是它的真假。

〔三〕第十节的数据、判定规则、编码与统计脚本可复算，取数时间为二〇二六年八月十日；判定规则在计算统计量之前写定，但取数后已先行查看过年份直方图，故本节按回溯检验报告，证据等级已在正文标明。

〔四〕第十节报告的 Fisher 精确检验为双侧，二乘二表由「存活是否满十年」与「是否带器材注记」交叉得到。R1 与 R2 不独立，正文已写明两者合起来只算一条证据。

〔五〕第十三节第十条提到的会计政策变更披露，本文未作任何超出其规范原文的推断；引用它是为了说明「基准变了要登记」在别处已是强制规范，而不是本文的发明。

〔六〕第十二节第三处的价值排序让步，是本文写作过程中被迫做出的最大一处删改。删去的原句为「每一次比较性判断都应当在记录中写明其另一侧」。保留这句话会与「不得把学生与具体同学并列写进档案」这条伦理要求冲突。

参考文献

- Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: a case study of the Basque Country. *American Economic Review*, 93(1), 113–132.
- Braun, H. I. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models*. Princeton: Educational Testing Service.
- Butterfield, H. (1931). *The Whig Interpretation of History*. London: G. Bell and Sons.
- Collingwood, R. G. (1946). *The Idea of History*. Oxford: Clarendon Press.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Espeland, W. N., & Sauder, M. (2016). *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, 18(8), 519–521.
- Hacking, I. (1986). Making up people. In T. C. Heller, M. Sosna, & D. E. Wellbery (Eds.), *Reconstructing Individualism* (pp. 222–236). Stanford: Stanford University Press.
- Hacking, I. (1995). The looping effects of human kinds. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal Cognition* (pp. 351–383). Oxford: Clarendon Press.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960.
- Hopkins, W. G. (2004). How to interpret changes in an athletic performance test. *Sportscience*, 8, 1–7.
- Hopkins, W. G., Hawley, J. A., & Burke, L. M. (1999). Design and analysis of research on sport performance enhancement. *Medicine & Science in Sports & Exercise*, 31(3), 472–485.
- Hyman, H. H. (1942). The psychology of status. *Archives of Psychology*, 269, 5–91.
- Merton, R. K., & Kitt, A. S. (1950). Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in Social Research* (pp. 40–105). Glencoe: Free Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Clarendon Press. (第四部分, 非同一性问题)

Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(3), 299–311.

Skinner, Q. (1969). Meaning and understanding in the history of ideas. *History and Theory*, 8(1), 3–53.

附 字数、版本与人机分工声明

正文约两万一千汉字（不含注释、参考文献与本声明）。版本：第一稿，二〇二六年八月十日。

人机分工：选题、三门学科的位置指派、判据的锐度要求与最终定稿由作者决定；文献检索、取数与统计脚本的编写与执行、初稿的组织由人工智能辅助完成。第十节的判定规则在计算统计量之前写入并存档；取数后曾先行查看年份直方图，此事已在正文与注释中两处标明。五条结果按原样写入正文，其中结果三直接改掉了本文原有的一句措辞（由「不留任何痕迹」改为「痕迹是角标不是字段」），结果四使本文放弃了对成绩差异的解释权，结果五承认两条判定规则不独立。

本文不提供任何自评分数。