

知觉的认证：一项关于中医脉诊能力制度性消解的认知-制度分析

——被删除的不是分数，是那个能力的存在资格

作者 胡敏 · SDE 学员 · 约 1.5 万字 · 发表于2026年7月29日 · 加固版

当前中医面临的合法性困境，通常被表述为一种认识论冲突——其核心概念（如气、血、经络）能否经受现代科学标准的检验。本文挑战了这一主流叙事，指出中医当代危机的真正根源不在科学检验本身，而在于过去三十年间，中医教育体系内部一套未经检视的资格认证技术格式，系统性地消除了从业者赖以产生这些概念的身体知觉能力。通过对执业医师资格考试制度、课程设置与临床教学评价体系的制度机制分析，本文揭示了以下核心发现：第一，中医教育评价制度内嵌着一个“命题性知识等同性假定”——即默认通过语词命题考试即可证明对一种深刻依赖身体知觉的认知体系的掌握，这在生成机制上构成了一个范畴错误。第二，这一假定在1990年代末至2000年代初因院校大规模扩招和统一考试制度建立的行政压力而被选中，并在随后的运行中，通过“教材-考题格式共构”“考官群体代际结构锁定”“被考核者身份认同绑定”三种制度根系，被锁定为几乎不可撼动的制度现实。第三，该机制导致了一个悖论性的“双重合法性幻觉”：外部评价体系误以为自己在评估一套保有操作传统的知识体系，被评估的从业者也误以为自己代表着这套传统，但实际上，评估所中介的核心能力——受训的身体知觉——已被考试技术从一代从业者的指腹下悄然删除。

本文的理论贡献在于，将评价制度的分析从“公平性”和“有效性”的常规议题，推进到“评价格式对认知能力的反身性消解”这一更深层级。我们识别出一个关键的区分，即“可被行为清单独立评分的操作能力”与“只能通过受训共同体的知觉恒常性来识别的能力”，并论证了对后者的评估，在本体论上不能依赖于独立第三方观察，而必须将评估者自身置于同一感知操作中。基于此，我们提出了一个替代性评价范式“操作者间知觉校准”及其可操作的2x2检验设计，该设计能干净地区分命题知识记忆与真实的知觉辨别能力，并给出了可能推翻本理论的具体的经验证伪条件。最后，本文提出了一个包含附加认证试点、数据积累与制度松绑的三轨递进式改革方案，为在维持既有制度框架稳定的前提下，系统性地修复中医教育的认知操作根基提供了路线图。

本篇收入「中医哲思」频道，并按频道规程做了一轮最近邻加固。「盲评」为原稿未经编辑改动时的分数；「打磨目标」是本轮的修改设计目标与编辑自评，不是认证分——按本站评分纪律，打磨者不得为自己改过的文本盖分，须独立评分者复核后方能作数。

1 引言

2019年，一项针对国内某知名中医药大学附属医院青年医师的匿名调查显示，超过40%的受访者在独立值班时选择依赖自动脉搏仪而非徒手脉诊来获取患者的心率与节律信息。他们并非不相信脉诊的价值，而是坦诚自己在临床上“摸不出来”。在一些病例讨论中，对“弦脉”或“滑脉”的描述，越来越像是对教科书定义的复诵，而非从指腹下直接涌出的知觉判断。

这个看似平淡的现象，揭示了一个远比“中医是不是科学”这类认识论争论更为迫切的危机。在学术界和公众舆论仍在为“经络能否被科学实验证实”之类的议题争论不休时，一个前提性的、更基础的消解过程正在静默地发生：需要通过活的身体知觉去捕捉的那个“对象”，正在新一代从业者的身体里系统性地丧失。这并不是某个具体的科学发现驳倒了中医理论，而是一项持续了三十年的制度实践，在经年累月、无人察觉的日常运作中，合法地、高效地、成功地使从业者不再具备生成和辨识其理论核心概念所需要的知觉能力。

本文旨在解剖这一制度性的消解机制。我们将论证，问题的症结并非外部科学共同体要求中医“证明自己”，这一外部压力固然构成背景，但真正致命的，是中医教育体系内部一个从未被显式追问的房角石假设——“命题性知识等同性假定”。这一假设为，如果一个知识可以以语词命题的形式被考试，那么掌握这个知识的充分必要证据，便是能在考试中正确回答相关命题。这个假定在面对“可被语言充分表征的知识”时，其危害是有限的。但当一个认知传统的核心对象——如脉诊中的“弦脉”与“细脉”的区分——其存在本身就依赖于操练有素的指腹在含混的身体感觉流中切分出稳定可辨的知觉模式时，将对其的评估完全托付给语词命题考试，无异于通过不断的制度筛选，系统性地宣判那些无法被命题格式化的认知操作能力为非法、不必需，并最终将其淘汰。

我们将通过回溯资格认证技术格式的历史选择、分析其自我锁定的三种制度根系、并精微解剖考试命题技术如何从“认知操作的代理”异化为“认知操作的替代品”，来完整地呈现这一机制。在此基础上，文章将引入外科住院医师培训、专业音乐教育等“反向案例”的制度比较，以论证问题的根源不在于“考试”这一形式本身，而在于特定历史条件下，一种垄断性的、排斥操作直接评价的认证格式被选中并固化为唯一的制度通道。

本文的核心理论贡献在于：我们识别出了“可被独立第三方观察的行为能力”与“必须将评估者置于同一感知操作中才能识别的受训知觉能力”这两种能力在评价方法上的本体论区分。基于此，我们提出并系统阐述了一个替代性评价范式——“操作者间知觉校准”——的内涵、制度转化路径与可证伪的实证检验设计。这一设计能够回答一个撼动现有评价地基的问题：当我们将一批在语词命题考试中获得顶尖成绩的考生置于真正的临床触诊场景，并要求他们与一位持有体感能力的同行对同一位患者的脉象做出独立判断时，他们之间的一致

性，是会印证考试成绩的权威，还是会彻底崩塌至一个令人不安的低水平？这个可检验的预测，构成了本研究经验内核。

2 文献综述

本文所讨论的议题，位于医学教育、认知科学和制度社会学的交叉地带。我们将在以下三个脉络中定位本研究的贡献。

2.1 中医科学化与客观化验证研究

关于中医概念的现代研究，一个最庞大而持久的主流范式是“客观化”与“科学化”验证。该路径的预设是：气、血、经络等古代概念，应是某种尚待发现的生理或物理实体的朴素指称，研究者的任务在于找到其现代科学对应物。在此思路下，大量研究致力于探寻经络的解剖结构、气的生物物理学特性、以及脉象与血流动力学参数之间的相关性（如费伦等人关于穴位结缔组织特性的研究，以及近年来采用激光多普勒和超声技术探讨脉象形成机理的工作）。这一范式在推进中医被现代医学界部分接受方面贡献显著，但其最根本的局限也日益显现：它面临着一个无法回避的“翻译折损”问题。将“弦脉”还原为一组血管硬度与压力波形的参数组合，看似实现了客观化，但这个参数组合与一位经验丰富的医师在手下把握到的那种整体性、包含细微质感的知觉事件，在信息的丰度和维度上是否存在质的差异？当研究者试图用物理量去“解释”知觉时，其解释逻辑实际上是用一种感官（视觉，通过仪器读数）替代了另一种感官（触觉与本体感觉的融合），而原先那个知觉事件中被替代的感官通道的独特信息，便在翻译过程中被不可逆地删除了。这导致了一个悖论：越是成功的客观化，离那个需要被解释的原初认知操作——医师的切脉诊病——可能越远。

2.2 制度与文化分析：知识的社会塑形

第二条路径转移了焦点，不再追问“气是什么”，而是追问“关于气的知识是如何被社会性地生产和维持的”。这一路径吸收了科学知识社会学和医学人类学的洞见。有学者将中医的现代困境描述为一种文化冲突，即两种不同“医疗世界观”或“身体观”之间的碰撞，中医面对西方生物医学的文化霸权，其合法性被系统性地贬低。另一些研究则从制度主义出发，聚焦于中医药大学的课程改革、执业医师法的出台、以及中药产业现代化等具体制度变革，分析其如何重塑了中医的身份与知识生产模式。这类研究指出，大规模的院校教育取代传统师承，使得蕴含在师徒亲密互动中的默会知识传递路径被切断；国家执业医师资格考试的建立，则将一套外在于中医传统的、侧重于命题性知识考核的标准化评价制度，确立为专业准入的唯一门槛。

上述分析有力地揭示了外部社会结构和文化权力对中医的塑形作用。然而，它们大多停留在一个“从外向内”的宏观叙事：即一个外部的、异质的评价标准（科学）被强加于一个内部的知识传统（中医）之上，导致了后者的扭曲与衰落。这种叙事无法有效解释以下关键问题：为什么这套被批评为“不合适”的评价标准，能够在中医教育体系内部如此根深蒂固地运行近三十年，以至于即便有改革的意愿，也难以撼动？它仅仅是因为“西方科学霸权”的外部压力，还是在中医教育体系内部，长出了自身的、独立于外部意志的维系力量？

2.3 具身知识与技能习得：从笛卡尔式裂缝之外看

第三条路径来自认知科学和技能习得研究，尤其是对“默会知识”和“具身认知”的探讨。波兰尼的“我们知道的比我们能言说的多”这一经典命题，为理解中医这类深度依赖身体知觉和情境判断的实践传统提供了有力的哲学基础。在医学教育领域，以OSATS（客观结构化评估工具）为代表的现代临床技能评价体系，正是试图弥合“知”与“行”之间的鸿沟，通过标准化的操作清单，对住院医师的临床操作进行可观察、可量化的评估。然而，我们对这一脉络的审视将揭示其一个隐含的边界：OSATS及其延伸框架，其成功应用几乎完全局限于那些“产生清晰、可被外部观察者独立记录的行为标记”的技能，如外科手术操作、心肺复苏流程。评估者作为独立第三方，即便对操作者的内心判断一无所知，仍可通过核对行为清单来完成评价。

在这些研究中，存在一个集体性的盲区。当面对另一种能力时，上述框架便陷入失语。这种能力，我们称之为“不产生外部可观察行为标记的受训知觉”。一位经验丰富的香料调香师能分辨数千种气味的细微差异，一位专业的品酒师能准确说出葡萄的品种与年份，一位中医师能通过指腹的微妙感觉区分28种脉象。他们的判断本身（“这是龙涎香”，“这是1982年的拉菲”，“这是弦脉”）就是能力的唯一展演，这个展演过程不必然伴随任何可被一个“无知”的第三方独立观察者核验的外部行为线索。此时，评估的逻辑发生了根本性转变：评估者不再是观察行为清单的审核员，而必须是另一个同样拥有该受训知觉能力的判断者。评估的通道，从“独立的看”转变为“共感的触摸”。中医脉诊技能正处于这一盲区的核心地带。

2.4 文献缺口与本研究定位

综上所述，现有研究或则尝试将中医知识“翻译”为普世科学语言（科学化路径），或则批判这种翻译背后的文化-制度霸权（文化/制度分析），或则提供了理解其认知特性的哲学框架，却无法为这类不产生外部行为标记的具身知觉能力提供一套有效的评价方法论（技能习得研究）。一个核心的断层就此显露：当中医教育试图在制度上证明其从业者的专业能力时，它别无选择，只能求助于现有的、为另一种类型的知识（命题性知识）和技能（产生外部行为标记的技能）而设计的认证技术。这一制度上的“被动选择”，其深远效应远未被充分揭示。

本文的定位，即在于填补这一断层。我们不再追问“中医知识是否科学”，也不仅仅批判“科学标准的不适用”，而是将分析推进到认证制度的具体技术格式这一中观层面。我们的核心论点是：中医当代危机的关键，不是“东方知识不被西方认知承认”，而是“不产生外部行为标记的受训知觉能力”被一套在制度上居于垄断地位、且在生成机制上无法对其进行有效辨识的资格认证技术格式，通过年复一年的考试、教学、和课程设置，从一代又一代从业者的身体里系统地消解了。问题的根源是认知-制度的，而非文化-政治的。下文将展开这一论证。

3 理论框架

3.1 核心概念定义

在进行机制分析前，有必要对几个核心概念给出操作化定义，以区别于直觉性理解。

命题性知识等同性假定：指在教育与评价制度的设计中，一个被普遍默认却很少被显式声明的假设：若一门知识体系的核心内容能够被表述为语词命题，则学习者是否掌握该内容，其充分必要的证据便是其能否在考试中正确回答由这些命题构成的试题。此假定省略了对另一类证据——即学习者能否在真实或高仿真情境中，稳定地执行该知识所对应的认知操作或身体技能——的制度化要求。我们将论证，此假定是当代中医教育困境的深层根源。

受训知觉与操作者间校准：

- 受训知觉：指一个感知系统（如触觉、嗅觉、听觉）经过长期的、有反馈的刻意训练后，能够在一个含混、高噪声的感觉流中，稳定地切分出特定模式，并赋予其共识性命名的能力。中医脉诊、香水晶鉴、音乐家辨别绝对音高，均属此类。
- 操作者间校准：评估受训知觉唯一有效的元方法。由于不存在一个独立于知觉主体的“金标准”（如测量血压的血压计），确认一个知觉事件（如“弦脉”）是否真实发生，只能依赖于两个或多个接受过同样训练、并被证明在历史中能进行稳定识别的知觉系统（即合格的操作者），对同一对象（同一位病人的右关脉）进行独立操作后，观察其判断是否趋于一致。评估的本质，从“独立观察者对行为的核查”转变为“被校准的知觉系统之间的互验”。

制度根系：指一套制度性安排（如一项考试规则、一种课程设置模式）在运行足够长时间后，会在其周围的社会环境中生长出的、使其自身难以被替换的三种结构性依存关系，详述如下。

3.2 命题与假设

基于以上定义，我们提出本研究的核心命题与工作假设：

命题1a（认知-制度错配命题）：中医核心诊断技能（脉诊）的认知本体，属于“不产生外部可观察行为标记的受训知觉”。评估此种能力的必要条件，是评估者本人具备同等水平的受训知觉，并能与被评估者进行“操作者间校准”。

命题1b（制度消解命题）：由于国家执业医师资格考试的垄断性地位及其采用的技术格式（以选择题和操作清单为主），它系统地排除了实施“操作者间校准”的可能性。其结果是，该考试持续、合法地筛除了对“受训知觉”能力的考核需求，从而在代际层面消解了此能力。

命题2（制度锁定命题）：命题1b描述的消解机制并非不可逆转，但其自我修正的尝试会受到三种“制度根系”——(1)教材-考题格式共构、(2)考官群体代际结构锁定、(3)被考核者身份认同绑定——的强力约束，使得改革面临极高的制度性交易成本。

核心可证伪假设（H1）：在控制了学历年限与既往考试平均成绩后，一群在中医执业医师笔试中成绩顶尖（如前10%）但从未经历过正式的师徒制脉诊校准训练的考生，在临床真实患者身上进行脉诊时，其与资深持证校准医师的判断一致性（Cohen's Kappa系数），将不会显著高于机遇水平（接近于零）。这一假设构成了本文最强的经验内核。如果结果相反——即高分笔试直接预测了高水平的临床知觉一致性——那么本文关于认知-制度错配整个理论地基将被动摇。

3.3 跨学科理论资源与适用性边界

本研究的分析，有选择地整合了来自三个领域的理论资源。

第一，来自科学知识社会学的“实践转向”，尤其是其对实验室研究中“技能”的非命题性传递的强调。这帮助我们理解，知识不仅是理论，更是熟练的身体技巧。然而，我们与其不同之处在于，不再将中医的困境简单归因于“西方科学的文化霸权”，而是将其置于不同制度技术的比较之下。制度压力是真实存在的，但它如何被内部化为一套自我消解的考试指令，才是分析的关键。

第二，历史制度主义，特别是其关于“路径依赖”和“正向反馈”的机制。该理论指出，在某个“关键节点”被选择的制度，会因其带来的收益递增效应——例如行政便利、利益集团的形成、认知惯性的锁定——而变得越来越难以偏离。我们用“制度根系”这一概念，

将正向反馈机制具体化为三种可识别的结构，从而可以清晰地分析中医资格认证制度的锁定机制，而不仅仅是抽象地感叹“改革很难”。

第三，感知心理物理学与专长研究，关于“知觉学习”的机制。有大量研究证实，人类感知系统经过训练，能够在复杂的物理刺激中发展出超乎常人的辨别能力（例如，放射科医生阅读X光片，品酒师的嗅觉辨别）。本研究借用其方法论内核——即通过测量受训者间的一致性来确证知觉事件的信号存在性——来构建我们的“操作者间校准”范式。但我们的应用边界十分清晰：我们不试图用心理物理学实验去验证“气”的物理实在，那将退回到文献综述中批评的客观化路径。我们只是论证，对于一个声称“我能摸出来”的从业者群体，最可靠的验证其声称之真的方式，就是测量他们彼此之间的知觉共识。如果这种共识系统性地崩塌，我们就必须认真对待一个判断：他们共同指称的那个知觉对象，可能已经不在他们指腹下的“经验世界”里了。

4 方法论

4.1 研究设计

本文是一项基于机制的认知-制度分析研究，其目的不是对一个现象的宏观定量描述或因果推断，而是打开导致该现象发生的制度“黑箱”，详细解剖其内部的运行齿轮与锁扣。研究方法上，这属于“机制解释”或“过程追踪”类型的研究。

选择此种方法，是因为我们试图回答一个“如何可能”的问题：中医的诊断技能，作为一种在两千年的实践历史中被反复验证和传递的活的经验，是如何在短短三十年的现代教育制度中被系统性地消解的？这个问题无法通过简单的相关性分析（如“扩招率”与“学生脉诊能力评分”的相关性）来回答，因为此类分析无法揭示制度运行的微观行为逻辑。我们必须拆解制度设计的多个环节——从政策文件、考试大纲、到教材编写、课堂实践——形成一个完整的逻辑链环。

4.2 分析策略与步骤

我们的分析将遵循一个四步推进的逻辑：

第一步：历史关键节点分析。我们将定位1990年代末至21世纪初这一时间段，论证何以一种特定的资格认证技术格式（以大规模、低成本、客观评分为导向的选择题与操作清单）被“选中”。我们关注的基础不是参与者的主观意图，而是当时面临的制度环境约束——包括高等教育大规模扩招政策、国家统一司法和医师考试制度的建立、以及国际医学教育认证标准化浪潮——三者如何历史性地耦合，使效率逻辑压倒了认知忠诚。

第二步：微观机制解剖。我们将下潜至“命题技术”这一颗粒度，精细地分析一道典型的“弦脉”执业医师考试题，是如何通过将感知事件压缩为语词定义，在年复一年的考试中，将“记忆语词定义”构建为等同于“掌握诊断能力”的制度性等价物。

第三步：反向案例比较。为了将我们的论点从“所有考试都会杀死操作技能”这一过于宽泛的悲观论断中剥离出来，我们将引入外科住院医师规范化培训和专业音乐学院演奏家培养，作为成功的反向案例。通过比较，我们将揭示，问题不在于“考试”这种形式，而在于是否在制度设计中，为不产生外部行为标记的受训知觉能力，保留了独立的、不被命题考试所替代的评估通道。

第四步：构造可证伪的检验设计。基于以上理论分析，我们将构建一个核心的可证伪假设（即前述H1），并设计一个具体的2x2研究设计。该设计能干净地区分“命题知识记忆”和“受训知觉能力”，并对我们的核心论点和主要竞争解释（“高分学生知觉能力自然更强”）构成直接的检验。

4.3 方法局限的预先承认

本研究作为一个以概念分析和机制理论建构为主的研究，其局限是显见的。第一，它缺乏大规模的经验数据来直接检验其核心假设H1。我们提供的，是基于理论推演的可证伪预测和一套完整的检验方案，但为其提供铁证，则有待后续的实证研究。第二，历史制度分析依赖对已发生记录的二次诠释，可能存在着过度强调结构必然性、而忽略历史中偶然因素和能动者作用的危险。第三，我们的比较案例分析（外科培训等）是选择性的，是为了理论建构的需要而选取的典型示例，其代表性可能存在偏差。

尽管如此，我们认为一个深刻的、可被证伪的理论，本身就是一项重要的学术贡献。它能够重新组织我们对看似熟悉的现象的理解，诊断出问题的症结，并为未来的实证探索和制度改革提供一个清晰的、可被检验的航标。

5 分析与讨论

本章将分四个部分展开核心论证。第一，我们将历史地追溯“选择题”如何成为中医评价的基石；第二，我们将解剖这种评价格式自我锁定的“制度根系”；第三，我们将通过反向案例比较，逼出关于“不可见能力”的核心范畴；第四，我们将系统性地提出并阐述“操作者间知觉校准”这一替代性范式及其检验设计。

5.1 效率的逻辑：选择题如何被“选中”

1999年，中国高等教育开启了影响深远的大扩招。在这一宏观政策背景下，全国各中医药院校的招生规模在短短数年间成倍增长，从精英教育迅速转入大众化教育阶段。几乎在同一时期，国家开始建立统一的执业医师考试制度，旨在对各类职业的从业者设定一个标准化的准入门槛。在医学领域，《中华人民共和国执业医师法》于1999年正式实施，中医执业医师资格考试随之建立。

这三股力量——招生规模扩张、统一考试制度的诞生、以及与国际医学教育标准化趋势的接轨压力——在同一历史节点上耦合，对中医教育的资格认证提出了一个全新的、极其刚性的行政需求：必须设计出一个能对数以万计的、来自不同教育背景的考生，进行快速、公正、低成本且能排除考官主观偏见的大规模统一筛选机制。

面对这一需求，选择题（及其变体多选题）和结构化的操作清单，几乎是唯一的行政解决方案。相较于一位老中医通过数周甚至数月与学徒的共同生活和临床带教来评判其能力，一场150分钟、300道选择题的考试，在行政效率上具有压倒性的优势。选择它，不是因为它在识别“好中医”的认知操作上更有效，而是因为它是唯一能在给定的人力、物力和时间成本下，实现大规模筛选功能的行政技术。在当时的语境下，甚至没有第二个选项被严肃地提上议程。

这一选择的行为本身并非恶意。初始的政策制定者或许设想，书面考试用以评估理论知识掌握，而临床实习和面试则可以用来把关实践操作能力。然而，这个设想的致命缺陷在于，它为两种评估方式赋予了完全不均等的行政合法性和可操作性。笔试成绩是刚性的、可量化的、易比较的；而基于临床观察的评估，尤其是在取消了传统师徒制、一个带教老师面对数十名学生的院校教育环境中，则迅速退化为一种主观的、形式的、几乎没有区分度的走过场。

由此，一道致命的等式在制度运行的实践中而非任何政策文件中被悄然确立：

在选择题考试中取得高分 = 掌握中医理论知识 = 具备中医临床能力。

需要被评估的那个核心认知操作——指腹下微妙而精密的知觉辨识——在这条等式的运转中，从一个“必要条件”蜕变为一个“最好能有，但无法考核，因此不予强求”的点缀品。选择题的格式，从此不仅是一种测量工具，它开始成为被测量内容本身的定义者。

5.1.1 一道“弦脉”考题的微观政治学

为了将上述宏观制度分析落地，让我们解剖一道具象化的、虚构但典型的执业医师考试真题。

题目：

下列哪一项最能准确描述“弦脉”的脉象特征？

- A. 脉来濡缓，浮而细软
- B. 脉来弦急，如按琴弦
- C. 脉来数急，如珠走盘
- D. 脉来沉细，应指无力

答案：B

在这道完美的标准化试题背后，究竟发生了一系列怎样的认知置换？

- 第一步，感知对象被词语符号取代。 真实的弦脉，是一种瞬间涌入指腹的、多层次混合的触觉-本体感觉复合体，包含着张力、硬度、振动频率、以及沿指腹不同区域分布的微妙压力梯度。这道题将这些丰富的、不可言尽的知觉信息，压缩为一个抽象的文化隐喻“如按琴弦”。

- 第二步，词语符号的辨认被定义为能力的证据。 题目要求的，不是考生去感知一个真实的脉象并做出判断，而是从其记忆库中调取教科书上关于这个隐喻的解释，并将其与四个词语选项进行模式匹配。

- 第三步，也是最重要的一步，是这个判断动作的合法性被垄断。 在这个制度语境下，唯一合法的证明“我知道什么是弦脉”的方式，就是在四个选项选出B。一个考生如果能在临床实践中稳定地、准确地、在患者手腕上摸出弦脉，但因为在考场上忘记了教科书的原话而错选了其他，他在这套制度下就是不合格的。反之，一个考生可能从未真正体会过“如按琴弦”是何感觉，但因其出色的记忆力选对了B，他便制度性地“掌握”了弦脉。

这一分析揭示出，考试技术并非中立的媒介，它的内在格式本身就是一种强有力的筛选器。它持续地、年复一年地奖励着一种特定的认知能力——对语词定义的记忆力，并系统性地惩罚或无视另一种认知能力——受训的身体知觉。当整个教育系统的荣誉、奖学金、保研资格、执业证书和专业晋升，都无一例外地锚定在这个语词定义能力上时，一个理性的、处于高度竞争压力下的学生，会将其有限的认知

资源和训练时间，投入到哪里？是去人声鼎沸的门诊部，耗费数百小时去捕捉指腹下那个微妙的“弦急”之感，还是坐在安静的图书馆里，将那本考纲指定教材上所有脉象的ABCD定义背得滚瓜烂熟？

制度的筛选机制，通过学生理性的自适应行为，完成了一个闭环，并启动了其自我实现的预言。

5.2 锁死的逻辑：三种制度根系的生长

如果上述的“选择题置代”仅仅是一种出错的政策，那么纠正它似乎只需要在下一轮课程改革的“换一套题”即可。但现实远比这复杂。一个运行了近三十年的制度，已经在其周遭的环境中，长出了使其自身极难被撼动的结构。我们在此精微解剖这三种制度根系。

5.2.1 教材-考题格式共构

执业医师资格考试，作为全国性的高利害性考试，对整个中医教育生态系统产生了一种巨大的回溯效应——即“考试指挥棒”效应。一旦考试的题型和考核点被确定下来，教材的编写、课程大纲的制定、教师的课堂教学、以及最关键的——学生的学习行为，都会全面围绕它展开反向适配。

出版社为了增强教材的市场竞争力，会主动按照考纲和历年考题知识点来组织内容，将知识切分成易于转化为选择题考查的“知识点”单元。教师们备课时，讲解的重心会不自觉地偏向“哪些知识点是必考、重点的”。学生们则更直接，他们从入学第一天起，甚至更早，就在通过各种渠道打听“哪些课是水分，考前背题库就能过”。

这种三方互锁的共构关系，意味着我们试图替换的，绝不仅仅是一张试卷的题型。要改变它，就必须同步改写市面上几十上百种教材的体例，改变数万名教师的备课习惯和知识结构，并扭转数百万学生的备考行为模式。这是一个牵动整个产业生态的巨大工程，其行政协调成本高到足以让任何一位理性的改革推动者望而却步。

5.2.2 考官群体代际结构锁定

这个根系触及了一个更隐蔽、也更残酷的困境。当前，那些坐在国家医学考试中心和各种医师资格审定委员会中，手握“定义何为合格中医师”裁量权的资深专家和考官们，他们自身是如何成为专家的？他们中的绝大部分人，正是过去二十年这套以命题知识为核心的考试系统所筛选、塑造和认证出来的优胜者。

他们的卓越，主要体现在理论的精深、SCI论文的发表、以及在过去一系列标准化考试中取得的辉煌成绩。他们的能力结构本身，就是被旧考试格式所深刻塑形的结果。我们不能假设，这批在学术上取得巨大成功的专家群体，自身都天然地具备了对“弦脉”与“细脉”的精微体感辨识力。更可能的情况是，他们中的许多人，其个人的职业发展路径，正是对其诗词记忆和逻辑分析能力而非身体知觉能力进行制度性奖励的一个绝佳范例。

问题就此陷入死结。要改革考试制度，将“操作者间知觉校准”这类需要评估者本人也具备体感能力的考核方式引入，就必须由现在的考官委员会来主导设计。然而，如果考官委员会的核心成员，其资质正建立在旧制度的合法性之上，并且他们个人可能也缺乏未经语词中介的、纯粹的体感辨识力，他们如何去设计一场评估别人体感能力的考试？

这不仅是个别考官的态度问题，这是一个结构性的认知能力不对称。他们不是“不愿”改，而是很可能在以他们为主导的制度框架下，已经长不出那个去识别新的、需要被评估的能力所需的元能力了。这项制度的成功产物，构成了这项制度自我更替的最大障碍。

5.2.3 被考核者身份认同锁定

第三个根系存在于整个行业的从业者群体心中。迄今为止，所有现存的、持有合法执照的中医师，都是通过现行这套考试制度获得其职业合法性的。这意味着，他们的职业身份、社会地位、以及全部过往的奋斗和其所获得的体制内荣誉，都紧密地绑定在这套制度的权威性上。

承认这套评价制度的根本设计存在一个深层的范畴错误——即它无法有效识别并长期消解了核心诊断技能——会产生一个毁灭性的外延推论：这等于间接承认，所有通过该制度获得认证的从业者，在某种根本意义上，是“有缺陷的”或“不合格的”。

这种对个人身份和职业生涯的潜在颠覆性后果，使得整个从业者群体，在客观上成为了抵抗激进改革的保守力量。这种抵抗，可能不是以公开反对的形式出现，而是以更普遍的形式呈现：“改革是可以的，但必须渐进”、“不能否定历史成绩”、“我们这些老中医当年也是这样考出来的，不也好好的吗？”这类话语，其背后的驱动力，正是对被自身所经历的制度所否定的恐惧。

这三种根系——产业生态的共构、权力群体的能力锁定、从业者的身份认同——共同构成了一道又深又宽的护城河，保卫着那套三十年前因行政便利而被仓促选中的选择题系统，使其免于任何实质性的挑战。

5.3 不可见的技能：来自反向案例的对照

至此，我们的论述可能带来一个过于宽泛的悲观推论：即所有依赖标准化考试的制度，都必然扼杀实践技能。为了驳斥这一点，并使我们的论点更加精确，我们引入两个成功的反向案例：现代外科住院医师培训和专业音乐学院的演奏家培养。

案例一：外科住院医师客观结构化考核（OSATS）

在外科医生的培养中，OSATS是国际公认的金标准。它的成功之处，在于将“一台成功的阑尾切除术”这个复杂的操作，分解为一系列独立的、可被第三方观察者观察和评分的关键步骤（行为标记）：是否进行术前安全核查、消毒铺巾的范围是否恰当、皮肤切开的位置和层次、组织分离的手法、止血是否彻底、缝合的针距和松紧度，等等。一个有经验的外科主任考官，可以静立一旁，仅凭观察，就能在清单上客观地逐项打分，而无需亲自上台操作。

这项制度之所以有效，是因为它评估的是一种产生外部可观察行为标记的操作能力。评估通道是视觉，评估者的技能水平（只要他懂手术）虽有助于做出更精细的判断，但其理论上的合法性基础在于观察行为的记录，而非他本人手掌的触觉。

案例二：专业音乐学院演奏考试

一名小提琴学生的毕业考试，是其独立在台上完成一首高难度曲目的演奏。评估通道是听觉，评估者是台下坐着的多位资深教授和演奏家。他们聆听，并基于他们对音准、节奏、音色、音乐处理等要素的专业判断，给出一个综合评分。这个过程虽然充满主观判断，但它的对象——声音——是一个可以被公开、可以被共同感知和记录的外部事件。不同评委之间的一致性（评判信度），也是可以被计算和研究的。

将这两个成功案例与中医脉诊考试进行系统比较，一个清晰的区分边界显现出来。在OSATS和音乐学院考试中，被评估的操作，其本身或其直接产物，是作为一个公共的、外显的感知事件存在的。它不依赖于评估者是否亲自操作。评估者作为“知情的外部观察者”，其判断是被评估对象的行为或产出所锚定的。

而在中医脉诊中，情况发生了根本性的翻转。一个中医师摸到“弦脉”，这个判断本身，除了他自己那个口头报告（“这是弦脉”），不附带任何可以被一个“无知”的第三方独立观察者进行核验的外部行为标记。评估者无法通过“看他的手指”或者“听他说的话”，来判定他指腹下是否真的发生了那个知觉事件。评估的唯一通道消失了。

此时此刻，评估的合法性基础，就从一个“独立的外部锚”，不可救药地转移到了“评估者内部的知觉世界”。要判断这个考生摸得对不对，唯一的办法，就是一个有经验的人，自己的手指也必须搭在同一条手腕的同一个位置上，去亲自感受，然后在校准自己与他人的知觉。

这就引出了我们的核心论断：中医资格认证制度的根本困境，并非“选择题不好”，而是在制度设计和演变中，原则性地缺失了“必须将评估者自身置于同一感知操作中”这一替代评估通道。这种缺失，在一个被外部行为标记化的操作领域（如外科手术）中并不致命，因为OSATS可以作为补偿。但在处理那种“不产生外部可观察行为标记的受训知觉”时，这一缺失便直接导致了整个评价体系的崩溃——它终于失去了与被评估对象之间的、所需的最基础的信息通道。

5.4 分离的维度与可检验的设计：推进“操作者间知觉校准”

至此，我们已完成了对问题的诊断。从本部分开始，我们转向建构和检验一套可能的应对方案。我们提出的不是一个具体的政策，而是一个名为“操作者间知觉校准”的替代性评价范式，并为其提供了一个可直接进行实证检验的严格设计。

5.4.1 分离“判断的一致性”与“判断者的被校准程度”

当前医学教育研究中占主导地位的评价方法论，如以OSATS为代表的客观结构化评估工具及音乐表演评估中的陪审团一致信度研究，其底层逻辑是在一个存在“金标准”或共识性优绩标准的框架内，测量判断的一致性。它的核心问题是：多个评分者，对于同一份操作或作品，给出的分数是否一致？这种一致性是评分工具有效性的证明。

而“操作者间知觉校准”评估的，是完全不同的对象。它评估的，不是判断，而是判断者本身。它问的是：这一群声称自己能“摸到”弦脉的人，他们的身体知觉系统自身，是否已经被训练到足够稳定、精准，以至于可以在一个含混的感觉流中，不约而同、不依赖暗示地，捕获到同一个模式，并赋予其同一个名字？此时，多个被校准的身体之间的“一致”，就是证据本身，是那个知觉事件（弦脉）作为“真实”而发生的唯一本体论证据。这不是测量工具的信度，而是被测量现象的存在论条件。

这两者看上去都是计算“一致性”，但在生成机制上，衡量的是完全不同的存在。前者依赖一个外在于评估者的“客观标尺”；后者中，一群被良好校准的评估者自身，就是那个唯一的“尺”。正是因为现代医学教育评价制度未能分辨这一区别，才导致其将中医脉诊的考核评价，不断地推向无法识别的深渊。

5.4.2 一个让占位框架预测相反的2x2检验设计

澄清了理论层面的区分。现在，是让它接受实证烽火的考验。我们提出一个具体的、可操作的2x2因子设计，能够干净地区分出“命题知识记忆”和“受训知觉能力”，并让本理论与现有的主流评价理论（高分即高能）和竞争解释（自选择精英）做出完全相反的预测。

假设：我们的核心假设（H1）是，命题考试成绩（如执业医师笔试）与受训知觉能力（通过操作者间校准测量）是分离的。高分笔试不能替代知觉校准训练，且缺少后者，前者在面对真实临床信号时会“崩塌”。

操作化：

- 因素A：命题考试成绩。我们将在某中医药大学选取两组应届毕业生：一组为“高分组”，即其在校期间所有涉及脉诊相关理论课程的平均成绩排名在年级前10%；另一组为“对照组”，平均成绩在年级后50%。

- 因素B：校准训练背景。我们进一步将上述两组学生，根据其是否在大学期间参加过由经我们校准过的高年资医师带领的、为期至少20学时的、系统性的“操作者间脉诊校准工作坊”，分为“经历校准训练”与“未经历校准训练”两类。该工作坊的核心是：两位受训者同时触摸同一位病人的同一脉位，独立做出判断，然后由带教老师揭示自己的判断，并在复盘讨论知觉差异，直至达成一致。

至此，我们得到一个2（高分 vs 控制）x 2（经校准 vs 未校准）的完全交叉设计。我们将招募40位临床患者，患有典型的、经专家组共识确认的弦脉或滑脉。所有四组受试学生将被要求独立、依序地对这40位患者进行脉诊并做出强制选择判断。

检验的核心：判别格G2

我们将关注目光聚焦在格2的考生身上：即“高分组 × 未经历校准训练组”。这是一群在我们现行制度下最优秀、最顶尖的学生，他们是奖学金得主，是未来的学术之星，是那套旧考试系统的完美产品。

竞争性预测如下：

- 基于现有主流评价逻辑的预测（如OSATS框架的泛化推论）：主流医学教育的普遍信念是，扎实的理论知识是高级临床判断的基础。因此，他们会预测，格2的顶尖学生在临床上的真实判断能力，即便未经特别校准，也应处于中等或中止水平。他们的一致性系数（如Fleiss' Kappa）应该有0.4-0.6，因为他们已被证明是最聪明、知识最扎实的一群人。

- 本研究的核心预测：我们预测，格2考生的表现将发生崩塌。他们之间的独立判断一致性，以及与标准组的一致性，将不会显著高于机遇水平。他们的顶尖成绩，是“语词定义记忆能力”的完美展演，但这项能力在面对真实的、不附带任何语词标签的、指腹下的混沌触觉流时，是完全无效的。一个从没被教导如何在同一只手腕上与老师的指感进行校准的学生，面对一个真实滑脉时的感觉，和一个刚从校门口被拉进来的普通人，可能没有差别。因为他所拥有的，只有一个可以正确选出B选项的大脑，和一双从未被校准过的、无法将感觉“切分”与“命名”的手。

这，就是检验的全部。这个实验设计能同时排除几个最强的竞争解释：

1. 排除“自选择精英”论：一个强大的反驳会说，“愿意参加那个校准工作坊的，本身就是对脉诊有热情和天赋的‘精英’。”但我们控制了这一点。格2正是那群在旧制度下最成功的精英。如果他们的触诊表现崩塌，就完美地证伪了“旧考试选出的精英自然具备更好体感能力”这一论断。

2. 排除“知识万能论”：这是对“命题性知识等同性假定”的终极检验。如果格2的结果崩塌，就等于为这个假定下达了实验性死刑：我们终于有数据证明，在那个考场上被证明的“优秀”，在需要动用真实身体知觉的诊室里，是不合格的。

这个结果一经做实，它的效力就不再局限于学术探讨。它将是一个改变政策议程的、不可被忽视的硬信号。

5.4.3 本理论的自我证伪条件

一个理论的力量，部分体现在它明确给出了能推翻自己的条件。我们的整个理论，建立在一个核心的实证断言之上：命题考试高分与受训知觉能力是可以被分离开的。因此，能推翻本理论的，不是找到某个师徒传承下来的名医脉诊能力确实强于他考试分数低的徒弟，这不构成挑战。真正可以推翻该理论的研究结果，是反面的：在格2的检验中，我们发现，高分考生的临床一致性表现同样非常出色。如果在真实研究中，我们发现，即便从未参加过我们定义的校准训练，那批前10%的尖子生，对40位病人的脉象，仍然做出了高度一致的、与标准组一致的判断。这证明，执业医师考试的训练，已经以某种我们尚未察觉的方式，同步培养了学生的受训知觉能力；或者“高分即高能”是普遍成立的。面对这样的结果，我们必须承认失败。我们将不得不撤回我们对现行考试系统消解认知能力的指控，并承认我们过度复杂化了问题，也过度悲观地低估了语词学习与知觉能力之间的正迁移。

给出这一条件，意味着我们将自己的核心论点，正式地推入了波普尔意义上的“风险地带”。它不再是不可证伪的哲学思辨，而是一个等待着被铁一般的数据检验的经验命题。

5.5 综合讨论：意义、局限与三轨干预策略

本研究在理论上的首要贡献，是揭示并发掘了教育评价格式对认知能力本身的反身性。我们超越了将中医困境归因于“科学霸权”或“文化范式冲突”的宏观叙事，深入到制度运行的毛细血管层面，精确地诊断出一个被历史偶然选中、并由三种制度根系锁死的认证技术格式，如何在日复一日的运行中，合法地、无人蓄意地，将一套复杂的身体操作传统，从它的未来传承者身上训练没了。

因此，我们提出的改革方向，也必须是基于制度机制的微创手术，而非整体性的范式推翻。在此，我们提出一个必须按次序推进的“三轨递进式干预策略”：

第一轨（第1-3年）：附加认证与制度空间的创造。

改革的第一步不是去挑战和修改风险极高的国家统一考试，而是在其旁边，创设一个平行的、非强制的“附加认证”体系。可以在国内几家顶尖的中医药大学率先建立“脉诊操作者校准附加认证中心”。学生可以自愿报名，接受为期数月、总计超过200人次的严格校准训练，并最终通过与资深校准医师的同时触诊一致性考试。通过者，将被授予一个独立于毕业证的、但具有高度专业声望的“附加认证”证书，并在其简历和教育档案中被显式标记。这一步的关键在于“不破旧，先立新”。它不对现有的制度结构构成威胁，但悄然为“长出新芽”创造了一个不受旧评价逻辑笼罩的制度空间。

预见的失败模式：此认证可能因“对就业无实际帮助”而应者寥寥。对抗方案是，主动与国内几家顶级的、注重诊疗水平的中医馆和私立医院合作，使其在招聘时，将“附加认证”作为优先录用的条件。这样，就为证书注入了市场价值，吸引理性学生参与。

第二轨（第3-5年）：数据积累与信号突破。

在首批附加认证学生毕业并进入临床实践多年之后，我们可以获得进行核心检验所需的数据。此时，启动一项严格的、多中心的、令主流医学界无法忽视的对照研究。研究的关键设计，即是我们提出的2x2因子检验，尤其要刻意在“理论成绩无差异”的条件下，测试“持有附加认证的从业者”与“仅持有常规执照的从业者”，在同时触诊中的一致性差异。要特别控制住“自选择”的替代解释（如认证者本身就更优秀）。当第一批铁一般的数据——尤其是指向格2崩塌的数据——被发表在主流医学教育期刊上时，它将构成一个改变制度议程的“信号突破”。

第三轨（第5-10年）：制度根系置换与正式双轴评价。

当数据和公共舆论都积累到一定程度后，改革的盖子终于可以被撬开。此时，可启动正式的、但渐进的国家考试制度改革。目标不是用操作考核替换笔试，而是建立一种类似外科培训的“双轴”评价结构。笔试轴继续承担理论知识筛选的功能，占70%的权重。而另外30%的权重，则专属于“操作校准评估”轴。这一评估由获得认证的考官（考官自身必须通过跨考官交叉校准）主持，在多个标准化病人站点进行，核心目标正是评估其不产生外部行为标记的受训知觉能力。由此，命题性知识等同性假定，这个三十年的无形房角石，终于在制度层面被正式拆卸。

当然，本研究也有着深刻的局限。我们的整个分析，聚焦于评价制度的认知反身性，对政府行政权力结构、高等教育经济成本分摊、以及更广阔的中西医学学术政治等宏观制约因素，做了有意的悬置。特别是策略中的“外部委员会否决权”等议题，在中国的教育行政体制中面临的实操困难，本文未做深度的政治可行性分析。此外，我们关于制度历史的重构，在很大程度上是基于逻辑推断的“理想类型”，真实历史过程无疑更加复杂和充满偶然。这些都有待于更为深入的政治学和历史学研究来补足。

6 结论

6.1 核心发现

面对“中医是不是科学”这一百年追问，本研究给出的诊断是：中医在当代中国的真正危机，不是科学检验的失败，而是制度性的认知能力消解。其核心发现可凝缩为以下论断：中医核心诊断技能（脉诊）在生成机制上属于一种“不产生外部行为标记的受训知觉”，而主导其中医师资格认定的技术格式（以选择题为核心的国家统一考试），在主体论上无法识别此种能力。这套格式在1990年代末因行政效率逻辑被选中，并在过去三十年间，通过“教材-考题格式共构”“考官群体代际结构锁定”和“从业者身份认同绑定”三重制度根系，被锁定为几乎不可撼动的现实，导致其在日复一日的运行中，系统地将受训知觉能力从一代从业者的指腹下合法地、悄然地删除了。

6.2 理论贡献

在理论层面，本文深化了对教育和认证制度功能的理解。我们指出，评价制度不仅仅是被动的测量工具，它具有强大的反身性，能重塑所评价的认知对象本身。更为重要的是，我们识别并澄清了两种能力的本体论区分：一种是通过独立第三方观察评估的、产生外部行为标记的操作能力；另一种则是只能通过将评估者置于同一感知操作的“操作者间知觉校准”范式来识别的、不产生外部标记的受训知觉能力。我们论证，在认识论上混淆二者，并工具性地用前者的评估格式来垄断对后者的评价，是中医教育乃至所有依赖身体默会知识的传承体系，在现代化过程中所面临的一个普遍性结构陷阱。

6.3 实践与政策含义

本研究为徘徊于“全面科学化”和“回归传统师承”之间的中医教育政策，指明了第三条道路：一条基于精密制度设计的认知修复之路。具体的行动方案包括：

1. 创设附加认证体系：秉持“不破旧、先立新”的原则，在顶尖院校设立由持有体感能力的资深医师主导的、高强度的“操作者间知觉校准”训练和附加认证项目，为其授予独立于普通执照的专业信誉。
2. 开展关键的对照实证研究：启动严格的2x2因子实验，特别是检验“命题考试高分但在知觉校准上未经训练”的从业者的临床一致性，以生成撼动现有评价逻辑基础的数据。
3. 渐进推进国家考试的双轴制改革：在数据和专业共识积累到一定程度后，推动国家执业医师考试在笔试之外，纳入独立的、由被校准考官操作的、占总分30%权重的操作知觉评估轴。

6.4 研究局限与未来方向

本研究虽构建了一个严谨的认知-制度分析框架，并提出了可检验的核心假设，但其局限亦需坦诚面对。首先，核心假设H1的实证检验尚未完成，其可行性需要一项预调查来评估现实中的受试者招募是否可行。其次，纵向追踪研究带来了沉重的伦理挑战，特别是当研究发现受试者的体感能力在常规制度环境中不可逆地退化时，研究者应如何处理“告知”与“被动的医源性伤害”之间的张力，本文未能给出圆满解决方案，这可能是后续研究需要优先解决的伦理难题。

未来研究可沿以下方向展开：

1. 执行核心检验：启动我们设计的2x2因子实证研究，以获取对“命题性知识等同性假定”进行裁决的关键数据。这将是这一领域从理论批判走向经验科学的最关键一步。
2. 开展跨领域的知觉认证比较研究：将本研究框架移植至其他依赖受训知觉且不产生外部行为标记的领域，如品酒师、高级调香师的资格认证、音乐表演专业中“音色美感”评价等，以检验“操作者间校准”范式是否是一个可被普遍操作的跨领域认知能力验证方法论。
3. 进行体制内制度变迁的政治可行性研究：对附加认证试点和三轨干预策略在中国的现实政治与行政环境中运作的政治可行性进行深入分析，特别是研究在现有政府架构下，外部专业委员会的“否决权”能如何被操作化和制度化。
4. 科技修复路径探索：研究是否可能借助高精度体感模拟和AI即时反馈系统，在未来创造出一种不再完全依赖稀缺的“活人”校准者的“非人类”知觉校准训练技术，从而为这项古老技能在智能时代的传承，开辟一条全新的道路。

中医脉诊的困境，并非孤例。它像一面被放大至极端的棱镜，折射出我们时代一个更普遍的、沉默的危机：在追求公平、高效、客观的现代制度化进程中，我们是否在不经意间，正在合法地、系统地将那些无法被命题化、无法被清单化、必须通过一个活的身体传递给另一个活的身体的知识与能力，从我们的文明传承中无声地删除？

参考文献

- Collins, H. (2010). Tacit and explicit knowledge. University of Chicago Press.
- Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuition and expertise in the era of the computer. Free Press.
- Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. Academic Emergency Medicine, 15(11), 988–994.
- Foucault, M. (1975). The birth of the clinic: An archaeology of medical perception. (A. M. Sheridan Smith, Trans.). Vintage Books.
- Frank, R. G., & Zeckhauser, R. J. (2018). The comparative institutional analysis of health care. Journal of Comparative Policy Analysis: Research and Practice, 20(1), 1-17.
- Gallagher, S. (2005). How the body shapes the mind. Oxford University Press.
- Good, B. J. (1994). Medicine, rationality, and experience: An anthropological perspective. Cambridge University Press.
- Hsu, E. (1999). The transmission of Chinese medicine. Cambridge University Press.
- Jütte, R. (2005). A history of the senses: From antiquity to cyberspace. Polity.
- Kaptchuk, T. J. (2000). The web that has no weaver: Understanding Chinese medicine. Contemporary Books.
- Kleinman, A. (1980). Patients and healers in the context of culture: An exploration of the borderland between anthropology, medicine, and psychiatry. University of California Press.
- Kuhn, T. S. (1962). The structure of scientific revolutions. University of Chicago Press.
- Kuriyama, S. (1999). The expressiveness of the body and the divergence of Greek and Chinese medicine. Zone Books.

- Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.
- Mahoney, J., & Thelen, K. (Eds.). (2010). Explaining institutional change: Ambiguity, agency, and power. Cambridge University Press.
- Martin, J. A., Regehr, G., Reznick, R., et al. (1997). Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery, 84(2), 273–278.
- Merleau-Ponty, M. (1945/2012). Phenomenology of perception. (D. A. Landes, Trans.). Routledge.
- Noë, A. (2004). Action in perception. MIT Press.
- Pierson, P. (2004). Politics in time: History, institutions, and social analysis. Princeton University Press.
- Polanyi, M. (1966). The tacit dimension. University of Chicago Press.
- Scheid, V. (2002). Chinese medicine in contemporary China: Plurality and synthesis. Duke University Press.
- Schön, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books.
- Unschuld, P. U. (1985). Medicine in China: A history of ideas. University of California Press.

致谢

作者感谢匿名审稿人提出的建设性意见，以及清华大学科学史系在本文修改过程中提供的学术支持。文责自负。

利益冲突声明

作者声明无任何利益冲突。

最近邻加固与核心命题的重写

本篇收入「中医哲思」频道前做了一轮加固：把参照语料主动往外扩，请出最可能吞掉本文核心概念、而原稿未曾对质的那位对手，与它划一条可裁决的分离线（两套理论对同一件事作出相反预测），再把核心命题从肯定判断改写为否定—重命名判断，并补上「本文禁止什么」。

一、原稿没有请出的那位对手：柯林斯的「交互型专长」与「贡献型专长」

本篇在文献里点了柯林斯的名，引了《默会与显性知识》，但没有正面对质他最要命的那一对区分：交互型专长（interactional expertise，能就一个领域流利对话、做出与内行难以区分的判断，却不会亲手做）与贡献型专长（contributory expertise，能亲手做）。柯林斯用「模仿游戏」实验证明：色盲者可以获得关于颜色的交互型专长，先天盲人则不能；而在很多领域，语言测试足以把内行与外行分开。

这一对区分正对着本篇的心脏。本篇主张语词命题考试无法评估知觉能力，是范畴错误；柯林斯却用实验说明，语言测试在很多情况下确实能检出深层的、身体性的领域浸润——因为交互型专长本身就是共同体的语言实践中被身体性地习得的，它不是空洞的复诵。

不请出他，本篇最容易被这样反驳：你说考试只测到了嘴上功夫，可柯林斯已经证明嘴上功夫本身就是一种真专长，而且和手上功夫高度相关。

二、分离线

分离线在于两种专长在被认证时是被分开还是被混同。柯林斯的框架预测：交互型与贡献型虽可分离，但在正常运作的共同体里两者高度相关，因为交互型专长是在与做者的长期语言浸泡中获得的——语言测试因此是贡献型专长的一个可用代理。

本篇主张：一旦认证格式垄断性地只走语词命题这一条通道，它就不再是代理，而成为定义——制度不再用语言去间接检测知觉，而是把语言表现直接写成了「具备该能力」的构成性条件。此时两种专长的相关性不是天然常数，而是被制度调节的变量，且会随格式垄断的年限单调下降。这是一个可被查错的历史命题，不是一句哲学断言。

相反预测，可做：把执业医师考试成绩与盲法脉诊辨识成绩（同一批标准化受检者、隔屏只允触脉、以操作者间知觉校准为评分口径）做 2×2 ，并按毕业年代分层。柯林斯框架：两项成绩在各代际中稳定正相关，语言测试可作贡献型专长的代理。本篇：整体相关微弱，且相关系数随毕业年代单调下降——近十五年队列显著低于更早队列；同时高语言分/低知觉分的「纯交互型」比例逐代上升。若各代际相关系数无差异，「认证格式反身性消解」即被推翻，本篇应退回为「考试效度不足」这一常规诊断。

三、核心命题的重写

资格考试不是没有测准那个知觉能力，而是把「能谈论知觉」重新定义成了「有知觉」——它删掉的不是某些人的分数，是那种能力

在制度中继续存在的资格；被淘汰者不知道自己失去了什么，因为衡量得失的那把尺子已经不再包含它。

四、本文禁止什么

禁止推广到可被行为清单独立评分的技能。准入条件是该能力只能通过受训共同体的知觉恒常性被识别；缝合、插管、影像判读这类可由第三方观察逐项打分的操作不在射程内。

禁止用它反对资格考试制度本身。本篇提出的是附加认证轨道与操作者间知觉校准，不主张废除现行考试。

禁止用它判定任何具体从业者的能力。本篇的分析单位是代际队列与制度格式，不是个人；用它去评价某位医师是「被清空的一代」属误用。

禁止把交互型专长贬为无用。本篇主张的是二者在认证中被混同，不主张能谈论者一无所长。

本轮加固所核验的文献

Collins, H., & Evans, R. (2007). *Rethinking Expertise*. University of Chicago Press. (交互型／贡献型专长，本轮所对质的核心对手)

Collins, H. (2004). Interactional expertise as a third kind of knowledge. *Phenomenology and the Cognitive Sciences*, 3(2), 125–143.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). Macmillan. (构念无关方差与构念代表性不足)

Biggs, J. (1995). Assumptions underlying new approaches to educational assessment. *Curriculum Forum*, 4(2), 1–22. (考试对教学的回冲效应)