

《心理科学进展》审稿意见与作者回应

题目：心理治疗的忠实度测量：现状与改进

作者：杨浩然、孙焕翔、刘晓明

第一轮

编委 1 意见：

由于心理治疗流派众多，从社会科学到自然科学各个领域都会涉及，有的强调实证科学，有的间于科学与艺术之间，这样宽广的领域来谈质量评估是非常难得事情。因为心理治疗质量涉及到科学与艺术的结合。鉴于此，建议作者更聚焦科学的、循证的心理治疗研究领域，针对临床明确问题的治疗，在这一领域展开更有意义。如果在学校、偏社会层面，问题性质不好确定，干预方案或方法千变万化，或者在临床实践中这样来进行保真度或质量评估难度太大。

回应：

非常感谢编委专家对稿件的细致审阅和提出的宝贵修改建议！我们十分认同您的观点，将忠实度测量置于循证心理治疗的框架下进行探讨，的确更有意义。为此，我们对返修稿的 6.3 节进行了针对性修订，具体说明如下：

首先，我们阐述了循证实践与忠实度测量之间的内在逻辑关联，论证了在循证心理治疗领域测量忠实度的合理性和必要性。借鉴 McLeod 等(2013)的研究框架，我们明确阐述了忠实度各成分与循证实践三大核心要素的对应关系：即，依从性体现了“最佳研究证据的实施”；技术胜任力对应“治疗师的临床专业技能”；而基础胜任力则涵盖了“对来访者特征、价值观与偏好的整合”。为夯实这一论述的理论基础，我们增补了 APA Presidential Task Force on Evidence-Based Practice(2006)、杨文登(2012)、杨文登等人(2017)及 Melchert 等人(2024)等关于心理学循证实践定义的相关文献。通过“对证据的追求”这一共同特点，将忠实度与循证实践建立关联。上述论述旨在强调，忠实度不仅是单一的治疗评估指标，更是连接实证研究证据与临床实践的桥梁。此外，我们进一步引用了 Margison 等人(2000)“测量是循证实践的基础”的核心观点，并结合 Bond 和 Drake(2020)、Cook 等人(2017)及 Marriott 等人(2025)的研究，论证了忠实度测量是循证心理治疗得以实施、维持与推广的前提条件，并将忠实度直接关联至临床服务质量的评估与持续改进(Goldman et al., 2001)。通过上述修订，我们将忠实度测量的讨论置于科学、循证的框架之内，这在现实研究和实践背景中是更符合实际的选择。

其次，针对您提出的“部分心理治疗流派兼具科学与艺术属性，跨流派质量评估难度较高”等问题，我们在返修稿中采用了更为审慎的表述。一方面，我们首先承认在特定治疗情境中，忠实度测量确实面临技术与情境的困难；另一方面，我们也补充了来自心理动力学(Ogrodniczuk & Piper, 1999)、人本主义(Freire et al., 2014)、心智化治疗(Karterud et al., 2013)及家庭治疗(Robbins et al., 2011)等多个心理治疗流派的忠实度评估量表，说明即便是这些更强调关系性与体验性的治疗取向，也已发展出相应的忠实度测量工具。这些证据表明，对心理治疗过程进行可量化的质量评估已成为行业共同需求。最后，如您所述，对于学校与社区等资源有限、问题性质复杂的环境，忠实度测量的实施门槛较高。因此，我们提出，应将忠实度测量的适用性与有效性作为未来实证研究的重点方向，通过开发更灵活的测量工具并进行严格的效度检验，逐步积累证据。

以上修改详见“6.3 从忠实度测量到心理治疗质量评估体系的建立”节，再次感谢您的耐心审阅！

参考文献：

杨文登. (2012). *循证心理治疗*. 北京: 商务印书馆.

杨文登, 李晓苗, 张小远. (2017). 心理治疗循证实践中“证据”的四个基本问题. *心理学报*, 49(6), 841.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271–285.

Bond, G. R., & Drake, R. E. (2020). Assessing the fidelity of evidence-based practices: History and current status of a standardized measurement methodology. *Administration and Policy in Mental Health and Mental Health Services Research*, 47(6), 874–884.

Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics*, 14(3), 537–545.

Freire, E., Elliott, R., & Westwell, G. (2014). Person-Centred and Experiential Psychotherapy Scale: Development and reliability of an adherence/competence measure for person-centred and experiential psychotherapies. *Counselling and Psychotherapy Research: Linking research with practice*, 14(3), 220–226.

Goldman, H. H., Ganju, V., Drake, R. E., Gorman, P., Hogan, M., Hyde, P. S., & Morgan, O. (2001). Policy implications for implementing evidence-based practices. *Psychiatric Services*, 52(12), 1591–1597.

Karterud, S., Pedersen, G., Engen, M., Johansen, M. S., Johansson, P. N., Schlüter, C., ... & Bateman, A. W. (2013). The MBT Adherence and Competence Scale (MBT-ACS): development, structure and reliability. *Psychotherapy Research*, 23(6), 705–717.

Marriott, B. R., Meyer, A. E., Feagans, A., Batch, B. L., & Adams, Z. W. (2025). Characterizing Fidelity Monitoring Practices in Community Behavioral Health Care. *The Journal of Behavioral Health Services & Research*, 1–20.

Melchert, T. P., Halfond, R. W., Hamdi, N. R., Bufka, L. F., Hollon, S. D., & Cuttler, M. J. (2024). Evidence-based practice in psychology: Context, guidelines, and action. *American Psychologist*, 79(6), 824.

Ogrodniczuk, J. S., & Piper, W. E. (1999). Measuring therapist technique in psychodynamic psychotherapies: Development and use of a new scale. *The Journal of psychotherapy practice and research*, 8(2), 142.

Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of consulting and clinical psychology*, 79(1), 43.

.....

审稿人 1 意见：

作者围绕“如何改进心理治疗保真度测量长期存在的“高成本”与“评估不一致性”的核心问题展开，说明了研究背景、综述目的和问题的重要性，涵盖了该领域的发展脉络和现状，逻辑条理性较好，参考文献规范，反映出研究前沿，使用表格突出关键问题，核心概念进行了界定，体现了作者的独特见解，对该领域研究具有参考价值。但是根据综述的宗旨，该文有些方面需要进行调整和修改。

总体回应：非常感谢审稿专家对本文的认可！接下来，我们将结合您的具体建议和综述宗旨的要求，对您提出的问题进行逐一回复。

意见 1：结构性问题：该文第 6 部分“研究展望”，在展望未来的研究方向后，提出一个综合性的评估系统构想，“图 1 本文构想的一个基于大语言模型技术的保真度测量系统”。这个测量系统是作者的构想，并不是综述前人的研究结果，是开创性的工作，需要实证研究加

以验证,这样呈现不太符合综述的宗旨,可以在前人研究基础上提出作者自己的观点和看法,概括总结自己的创新。需要调整结构和表述;

回应:

非常感谢审稿专家指出的结构性问题,这对提升本文的严谨性十分重要。

我们完全认同您的观点,作为一篇综述,不应在结尾突然呈现一个看似独立的“系统”,让读者误以为这是一个已经完成的产品。事实上,本文 6.3 节是基于以往忠实度测量的“成本”与“一致性”痛点,结合大语言模型技术特性提出的测量方案。这个方案结合了我们自身的思考,同时也请教了心理治疗、精神医学、计算机等领域的相关专家,目的是为现有的忠实度测量提供一定的借鉴。

在返修稿中,我们对第 6 部分进行了如下调整:

(1)调整定位。我们将该部分内容的侧重点从“展示一个系统”调整为“探讨解决问题的方案”。我们明确指出,图 1 展示的框架是将前文综述中提到的“标准化框架及知识库(解决不一致问题)”与“大语言模型技术(解决高成本问题、部分解决不一致问题)”进行整合的一种方案,旨在为后续实证研究提供假设和方向。

(2)强化了内容的前后联系。在描述该方案时,我们增加了对前文综述内容的呼应,强调该方案是基于现有测量困境思考出的解决方案,而非凭空提出的新工作。

(3)修改表述语气。我们弱化了原稿中肯定式的表述方式,调整为探讨式的阐述,以符合综述类文章的写作规范。

修改后的稿件,在“改进措施”(本文第 5 节)、“未来探索方向”(本文 6.1 节)和构想方案的组成成分之间实现了较为一致的对应:

附表 1 各节论述结构的对应情况

现有改进措施	研究展望	测量方案
5.1 从概念到实施,建立忠实度的测量框架	6.1.1 持续完善忠实度测量框架	6.2 节:阶段(1)框架设计
5.2 从分类到体系,对干预措施进行精确描述	6.1.2 建立忠实度统一知识库	6.2 节:阶段(2)知识库建设
5.3 从人工到自动,使用计算机技术辅助评估	6.1.3 使用大语言模型技术辅助忠实度测量	6.2 节:阶段(3)测量实施
——	——	6.2 节:阶段(4)平台建立(未来)

具体修改内容详见“6.2 一个基于大语言模型技术的忠实度测量方案”。再次感谢您的耐心审阅!

意见 2:摘要:需要精炼和结构化;

回应:

非常感谢审稿专家的建议!原摘要的确略显臃肿,我们对此进行了结构化的精炼。

为突出摘要的实质性内容,修改后的摘要按逻辑将句子组织为:(1)研究背景(概念与重要性);(2)问题提出;(3)现状梳理(已有研究方法与不足);(4)研究构想;(5)研究意义。共五个句子。

修订后的内容为:“心理治疗的忠实度(fidelity)指心理治疗按预定方案实施的程度,是保障临床实践质量与验证研究效果的关键过程性指标。然而,受限于长期存在的“高成本”与“评估不一致”问题,忠实度测量的推广面临阻碍。本文系统梳理了忠实度的概念构成与测量必要性,比较了观察性编码、书面调查法与行为排练等常用测量方法的优势与局限,并评述了忠实度框架、干预措施精确描述、计算机辅助评估等现有改进思路的贡献与不足。在此基础

上，提出了一个基于大语言模型技术的忠实度测量方案构想，涵盖从评估框架设计、统一知识库建设，到自动化测量实施与共享平台构建的全流程。本文旨在为国内心理治疗的过程质量评估提供研究思路与新视角，推动忠实度测量从理论倡议走向标准化实践。”

意见 3：关键术语的翻译：competence 建议翻译为：胜任力；common competence，建议翻译为基础胜任力。

回应：

非常感谢审稿专家对术语的建议。我们参考您的建议和国内学界的翻译习惯，已将 competence 译为“**胜任力**”、common competence 译为“**基础胜任力**”。此外，将 technical competence 相应译为“**技术胜任力**”。全文中涉及上述三个术语的地方均已更新，并用红色字体进行了标注。

意见 4：整体上是否强调循证的心理治疗理论流派比较重视心理治疗的保真度，突出心理治疗循证理念的重要性。“成熟的心理疗法通常配有经过验证的保真度评估工具”，这样的描述可能欠妥，如何界定“成熟”？

回应：

非常感谢审稿专家的建议和指正！关于“循证理念”和“忠实度”的辩证关系的确需要被进一步讨论，我们在返修稿的 6.3 节中对二者的关系进行了补充论证。此外，关于“成熟”描述的问题，已在稿件中进行了修改。

关于忠实度测量和循证心理治疗的关系，论证逻辑如下：首先，二者从概念上均对“证据”十分重视，在目标上亦有相似之处(即确保治疗按照要求实施)。依从性对应“最佳研究证据的实施”；技术胜任力对应“治疗师的临床专业技能”；基础胜任力则对应“对来访者特征、价值观与偏好的整合”(McLeod et al., 2013)。测量是循证实践的基础，测量程序的进步应扩展到循证心理治疗(Margison et al., 2000)。因此，循证需要忠实度的测量进行支撑。通过这一论述，我们强调了在循证实践框架下，忠实度测量不仅是评估手段，更是判断治疗师是否真正践行循证理念的关键前提(Bond & Drake, 2020; Marriott et al., 2025; Cook et al., 2017)。其次，即便是在传统上更强调关系性与体验性过程的治疗流派(如心理动力学)，出于对过程质量控制的共同需求，也正在积极开发相应的忠实度评估工具(如 Ogrodniczuk & Piper, 1999)。因此，未来研究需在多种心理治疗疗法及更复杂的实施环境中进一步探索忠实度测量的适用性与有效性。当然，我们的论述需尽量保持克制，将重点放在“过程质量监测”本身，避免文章重心偏移至“循证与非循证心理治疗的取舍和争论”上。以上修改内容详见“6.3 从忠实度测量到心理治疗质量评估体系的建立”，敬请您审阅。

此外，原文中“成熟的心理疗法”确实不够严谨。在返修稿中，我们删除了“成熟”表述，仅以举例的形式呈现典型的忠实度测量工具，使表述更符合学术规范，避免主观判断。具体修改为：“部分心理治疗流派已有经过验证的忠实度评估工具，如认知行为疗法的忠实度评估量表(Cognitive Therapy Scale-Revised, CTS-R) (Blackburn et al., 2001)、动机式访谈疗法的忠实度评估量表(Motivational Interviewing Treatment Integrity, MITI) (Moyers et al., 2016) 等。”

参考文献：

- Bond, G. R., & Drake, R. E. (2020). Assessing the fidelity of evidence-based practices: History and current status of a standardized measurement methodology. *Administration and Policy in Mental Health and Mental Health Services Research*, 47(6), 874–884.
- Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics*, 14(3), 537–545.

Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J. M., Audin, K., & Connell, J. (2000). Measurement and psychotherapy. *British Journal of Psychiatry*, 177(2), 123–130.

Marriott, B. R., Meyer, A. E., Feagans, A., Batch, B. L., & Adams, Z. W. (2025). Characterizing Fidelity Monitoring Practices in Community Behavioral Health Care. *The Journal of Behavioral Health Services & Research*, 1–20.

McLeod, B. D., Southam-Gerow, M. A., Tully, C. B., Rodríguez, A., & Smith, M. M. (2013). Making a case for treatment integrity as a psychosocial treatment quality indicator for youth mental health care. *Clinical Psychology: Science and Practice*, 20(1), 14–32.

Ogrodniczuk, J. S., & Piper, W. E. (1999). Measuring therapist technique in psychodynamic psychotherapies: Development and use of a new scale. *The Journal of psychotherapy practice and research*, 8(2), 142.

意见 5：总体上翻译的痕迹比较明显，建议从专业角度调整表述。比如：“传统机器学习方法因训练成本高、泛化性差而受限”。

回应：

非常感谢审稿专家细致的建议！我们为语言问题所带来的困惑致以歉意。

参考您的意见，我们认为是部分表述在语言流畅性有所不足。由于忠实度在国内研究中尚属空白，在查阅英文文献时，我们受到了英文表述习惯的影响，省略了对一些难懂术语的解释和将英语思维转化为汉语思维的过程，忽视了中文读者的阅读习惯。

针对您的意见，我们已对全文语言进行了润色，力求表述更加地道、流畅。例如，您提出的例子“传统机器学习方法因训练成本高、泛化性差而受限”，在返修文章中我们优化为“**传统机器学习方法存在高训练成本和低泛化性等弊端**”。类似的例子还如“因而心理治疗忠实度就成为效果研究的核心方法论考量(McHugh et al., 2009)”修改为“**因此心理治疗忠实度便成为疗效研究在方法论层面的核心考察指标(McHugh et al., 2009)**”。

全文修改之处用**红色**字体进行标注，请您审阅。

审稿人 2 意见：

论文对心理治疗中保真度的测量问题进行了系统梳理，指出了存在的问题，并提出用大语言模型来解决的思路，文献充分，写作规范，几个问题供参考。

总体回应：非常感谢审稿专家对本文的认可！也十分感谢您针对本文提出的建议。接下来，我们将对您关心的问题进行一一回应。

意见 1：treatment fidelity 是否可以翻译成治疗忠实度？和相关术语更接近一些，保真度多用于通讯工程、信号处理领域。

回应：

感谢审稿专家对术语翻译的宝贵建议。我们完全认同您的观点，“**忠实度**”的确是比“**保真度**”更贴切的译法。相关改动已在全文更新，并以**红色**字体标示。具体解释如下：

首先，从词源上看，韦氏词典对“fidelity”的核心释义是“忠实的品质或状态”(the quality or state of being faithful)和“细节的精确性”(accuracy in details)。因此，将 fidelity 直译为“忠实度”能够精准地表达词汇本意。其次，在相关研究中，“忠实度”这一表述也得到了应用。例如，在香港版动机式访谈法的评定工具手册中，就将“实际实施的治疗与预定治疗方案一致的程度”明确地称为忠实度(江嘉伟 等, 2025)。尽管该手册选用了近义词“Integrity”，但其概念与“fidelity”是相通的(Perepletchikova, 2011)。类似的，在医疗护理和教学实验等领域，学界也常采用“忠实度”来指代“对既定方案的如实执行程度”(李亨 等, 2019; 庞维国, 2011)。这

表明,“忠实度”符合本文语境的表达。

已有部分国内学者将“fidelity”译为“保真度”(艾小青 等, 2012; 李晓苗, 李占江, 2014; 张静 等, 2020)。然而,正如您所指出的,“保真度”一词在中文语境中,尤其是在通讯工程、信号处理及音视频等领域,已经形成了稳定的固定搭配(如“高保真”(High Fidelity, Hi-Fi))。若在心理治疗领域继续使用,可能会导致读者产生不必要的混淆,误解其核心含义,不利于概念的传播和推广。

综上,为了确保学术表达的严谨与清晰,我们最终将“**心理治疗忠实度**”作为“**psychotherapy fidelity**”的译名。同时,为了尊重已有文献的翻译习惯并为读者提供更全面的背景信息,我们在文中首次出现该术语时,采用了“**忠实度(fidelity, 亦译作保真度)**”的表述方式,以供读者参考。

参考依据:

1. 韦氏词典对 fidelity 词条(<https://www.merriam-webster.com/dictionary/fidelity>)的解释:

fidelity: 1a 忠实的品质或状态; 1b 细节的精确性; 2 电子设备或软件应用程序精确再现效果(例如声音或图像)的程度。

2. 翻译为“忠实度”的依据:

江嘉伟, 辛挺翔, 高蔚南, 许润好, 刘兆锋(译). (2025). 动机式访谈治疗忠实度编码手册(MITI 4.2.1 普通话版) (https://www.cmic-mi.com/wp-content/uploads/2026/01/miti_mandarin_4_2_1.pdf). 中华动机式访谈法学会。

李宇, 万巧琴, 于放. (2019). 干预性研究中干预忠实度的提出与发展. *中华护理教育*, 16(12), 945-949.

庞维国. (2010). 教学实验中的处理忠实度. *心理发展与教育*, 26(03), 322-328.

Francheska Perepletchikova. (2011). On the topic of treatment integrity. *Clinical Psychology Science and Practice*, 18(2), 148-153.

3. 翻译为“保真度”的依据:

艾小青, 曹玉萍, 张亚林. (2012). 内容分析方法在心理治疗研究中的应用. *中国心理卫生杂志*, 26(5), 373-376.

李晓苗, 李占江. (2014). 心理治疗疗效研究的方法学述评. *中国心理卫生杂志*, 28(3), 163-167.

张静, 陶红霞, 阮海慧, 牛晓丹, 迟俊婷, 王艳红. (2020). 动机式访谈法培训师的培养和实施保真度评估现状. *中华护理教育*, 17(12), 1134-1138.

意见 2: 保真度是针对于特定的操作而言, 想要构建一个统一的标准和测量是否可行, 似乎只是一个概括水平的问题。因此, 论文一方面指出保真度缺乏明确的评估标准和可操作的测量方法, 但在研究展望部分好像也缺乏可行的解决方案。

回应:

感谢审稿专家的建议! 这反映出我们需要重新审视原稿中的论述, 在“研究展望”部分必须更明确地将每一个忠实度测量的未来探索方向与具备可行性的解决方案进行对应。针对您的建议, 我们在修订稿的第 6 节(研究展望)进行了大幅重写与扩充, 具体如下。

我们在修改后 6.1 节每一个分节(6.1.1 节、6.1.2 节、6.1.3 节)的内容中, 阐述了这样做的目标、理由、可行方法, 在此基础上总结出一个可行的完整方案。在提出方案后, 我们在“**表 4 忠实度测量方案建设的改进方向**”中将完成忠实度测量的步骤、解决的问题、已有方法(以及未来可行的方法)进行了总结。具体叙述如下:

在修订稿 **6.1.1 节**中, 我们阐述了“测什么”的规范性问题, 指出应持续完善忠实度测量框架。需承认, 不同心理治疗疗法的操作存在差异, 因此我们需要的是一个能够包容这些**差异的框架**。在这一节中, 我们提出忠实度框架需借鉴 CONSORT(Hopewell et al., 2025)的更新思路, 结合 BCTTv1 分类法(Michie et al., 2013), 明确“依从性”和“胜任力”的通用描述方

法。尽管不同疗法的具体内容各不相同，但需为忠实度核心测量成分(依从性、胜任力)和各成分的细分维度(依从性的治疗成分、其他维度；胜任力的基础胜任力、技术胜任力)提供统一的测量框架，确保测量的结构一致。

在修订稿 6.1.2 节中，我们指出仅有框架是不够的，应建立包含具体治疗技术和实践方式的忠实度测量知识库，解决“如何界定特定操作”的标准化问题。为此，我们引入了 Leijten 等人(2021)的 7 种治疗成分提取策略，以**确保从特定流派中科学地提取核心成分**；此外，引入医学领域的“证据到决策”框架(Alonso-Coello et al., 2016)，**利用专家共识将分散的临床经验转化为具体的、可观测的行为锚点**。通过这一环节，我们期望未来的知识库能将每个治疗流派特异性的操作标准化为“可编码的知识条目”，调和通用标准与特定操作之间的矛盾。

在修订稿 6.1.3 节中，我们探索了“**如何实施测量**”的技术问题。即使有了框架和知识库标准，人工评估的高昂成本和评估不一致也是忠实度测量可行性的极大障碍。在本节，我们**论证了利用大语言模型进行自动化评估的可行性**。大语言模型擅长处理语言和语义分类任务，能够基于前述的框架和知识库对治疗文本进行高效、低成本的规模化评估，从而使忠实度测量走出实验室，具备了临床普及的现实基础。

在上述分项论证的基础上，我们在 6.2 节提出了一个整合性的实施方案，并绘制了流程图(见本文图 1)。同时，**最新修订的“表 4 忠实度测量方案建设的改进方向”详细列举了每一个步骤所对应的具体问题、已有的成熟方法以及未来的改进路径**。框架、知识库解决的是测量的可操作性，LLM 解决的是高效测量的可行性，三个阶段按照时间先后顺序构成了一个有反馈的流程，使得构建一个既能涵盖心理治疗特异性，又具备测量学统一性的评估体系成为可能。

此外，我们在 6.3 节中针对不同流派的特定干预操作，进一步展望了未来的探索方向。我们认为，尽管心理治疗的形式多种多样，但建立标准化的忠实度测量机制已成为心理治疗行业发展的共同趋势。未来的研究重心应转向学校、社区等资源有限且问题复杂的环境，通过实证研究进一步探索忠实度测量的适用范围与有效性。

再次感谢您的建议，帮助我们理清了文章的逻辑，使研究展望部分更加务实、具体。

参考文献：

- Alonso-Coello, P., Schünemann, H. J., Moher, J., Brignardello-Petersen, R., Akl, E. A., Davoli, M., ... & GRADE Working Group. (2016). GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*, 353.
- Hopewell, S., Chan, A.-W., Collins, G. S., Hróbjartsson, A., Moher, D., Schulz, K. F., Tunn, R., Aggarwal, R., Berkwits, M., Berlin, J. A., Bhandari, N., Butcher, N. J., Campbell, M. K., Chidebe, R. C. W., Elbourne, D., Farmer, A., Fergusson, D. A., Golub, R. M., Goodman, S. N., ... Boutron, I. (2025). CONSORT 2025 statement: Updated guideline for reporting randomised trials. *BMJ*, 389, e081123.
- Leijten, P., Weisz, J. R., & Gardner, F. (2021). Research strategies to discern active psychological therapy components: A scoping review. *Clinical Psychological Science*, 9(3), 307–322.
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions. *Annals of Behavioral Medicine*, 46(1), 81–95.

意见 3：论文提出构建基于大语言模型技术的保真度测量系统，但大语言模型似乎无法解决保真度存在的评估框架、评估标准等不一致问题，大语言模型只能解决成本问题。

回应：

感谢审稿专家的建议！我们非常认同您的观点，大语言模型(Large Language Model, LLM)

技术本质上是测量的实施执行工具，而非标准的“制定者”，其确实无法消除心理治疗领域长期存在的理论框架与评估标准的不一致性。这反映出我们在阐述基于 LLM 的忠实度测量方案时，应更充分地说明使用 LLM 的理由，以及客观地分析 LLM 技术可以解决忠实度评估的哪些问题。

针对您的建议，我们重新梳理了文章的逻辑结构，在修订稿的 6.2 节及图 1 中，明确界定了 LLM 技术在整个方案中的边界和功能，并对此做出了如下澄清与回应：

首先，我们在修订后的方案中明确将忠实度测量流程划分为两个独立部分：**(1)第一部分：通过“框架设计”(6.1.1 节)和“知识库建设”(6.1.2 节)，利用框架和专家共识来解决“定义层面的不一致性”。这是 LLM 无法替代的基础工作。(2)第二部分：只有在前述标准确立后，LLM 才介入“测量实施”(6.1.3 节)，将其作为结构化的提示词进行自动化评估。**

其次，阐述 LLM 在解决“不一致性”中的作用。虽然 LLM 不能制定标准，但我们认为它在解决“应用层面的不一致性”(即评分者信度问题)上具有独特优势，而不仅仅是降低成本。传统人工评估中，即使经过严格培训，不同评分者对同一标准的理解和执行仍存在差异。LLM 通过固定的提示词工程，能够严格遵循知识库中的定义和行为锚点，从而在执行层面实现超越人类的不确定性。针对 LLM 本身可能存在的随机性问题，我们在 6.1.3 节中补充了从原理(统计规律)、任务限定(规范化输入输出)到技术手段(结构化输出)的保障机制，论证了 LLM 输出结果的可靠性。此外，我们在文中补充了最新的实证证据(如 Wynia et al., 2025)，表明在给定清晰标准的前提下，LLM 的评估表现具有潜力和效度。修订后的图 1 展示了“专家制定标准”到“LLM 执行标准”的逻辑，表 4 则详细列出了各个阶段旨在解决的具体问题。

意见 4：一些语句的含义不太清晰，例如，“设计测量流程时，不同的研究人员很可能会根据资源的可用性而选择不同方法，或按照自己的意愿重复“造轮子”(Waltz et al., 1993)”。

回应：

非常感谢审稿专家的细致建议！我们为语言含义问题所带来的困惑致以歉意。

由于心理治疗忠实度研究在国内尚属空白，本文很大程度上依赖于英文文献。受此影响，我们在写作过程中未能完全摆脱英文表述习惯的束缚，导致部分语句存在直译生硬或使用非中文学术惯用语的问题，增加了读者的阅读负担。

针对您指出的具体例句，“可用性”(availability)这种表述在国内学界确实不太常见，重复“造轮子”(reinventing the wheel)指的是在已有相关工具的情况下，又推倒重来，开发相似工具的现象。为此，我们已在修订稿中进行了重写，修改后的例句为“**在缺乏统一标准的情况下，研究人员往往受限于实际资源而采用不同的测量方法，或忽视已有工具进行重复的开发(Waltz et al., 1993)**”。

以此为标准，我们对全文进行了语言润色与排查，对类似的英文比喻全部进行了本土化解释或直接删除。修改之处已在原文中用**红色字体**标注。

最后，再次感谢审稿专家提出的重要建议，相信修改后的稿件能更好地符合您的要求。

第二轮

审稿人 1 意见：

作者对提出的意见和建议进行了仔细的审阅和认真的修改，几个部分都得到了改善。但是仍然需要进一步修改。具体如下：

总体回应：

再次由衷感谢审稿专家对本文的细致审阅与宝贵批注！

为便于您和编委专家更清晰地审阅，同时向读者更直观地呈现所修改的具体问题，我们下文中将您批注时引用的原文表述以“灰色”字体呈现，并明确了批注的插入位置(例如：**【批注位置：第一轮审稿意见-审稿人 1-回应 1】**)。您批注的实质性建议则以**【意见】**标示。

具体修改及回应如下：

意见 1：【批注位置：第一轮审稿意见-审稿人 1-回应 1】“本文 6.3 节是基于以往忠实度测量的“成本”与“一致性”痛点，结合大语言模型技术特性提出的测量方案。这个方案结合了我们自身的思考，同时也请教了心理治疗、精神医学、计算机等领域的相关专家，目的是为现有的忠实度测量提供一定的借鉴。”**【意见】**作为综述，未来展望只是展望，简要提出未来可能的研究方向即可。目前这个部分所占比例太大，不太符合一般性综述的框架。

回应：

再次感谢审稿专家的建议！我们重新审视了本文的研究展望部分，目前版本的确包含过多原创性工作的探讨，偏离了综述的常规框架。

为此，我们作了较大幅度的精简和修改。具体而言，删除了原“6.2 一个基于大语言模型技术的忠实度测量方案”节整节内容(包括原流程图“图 1 基于大语言模型技术的治疗忠实度测量流程示意图”和表格“表 4 忠实度测量方案建设的改进方向”)。此项修改的具体细节与下文“回应 3”一致，相关内容已在正文“6 研究展望”部分以黄色字体进行了更新，敬请审阅。

意见 2：【批注位置：第一轮审稿意见-审稿人 1-回应 2】“(4)研究构想；(5)研究意义。”**【意见】**不太符合一般性综述的结构。一般性综述的摘要结构：背景/目的(背景介绍了，目的不太明确)，方法(交代作了什么，但是没有说怎么作的)，结果(作了各种方法的比较和分析，最好能够提炼出核心的发现)，结论(在前面结果的基础上，简要提出未来的方向)。

【批注位置：第一轮审稿意见-审稿人 1-回应 2】“在此基础上，提出了一个基于大语言模型技术的忠实度测量方案构想，涵盖从评估框架设计、统一知识库建设，到自动化测量实施与共享平台构建的全流程。”**【意见】**对应上述摘要 5 个部分的哪一个部分？是研究构想？作为综述来说，研究构想占的篇幅过大不太合适，感觉作者还是没有体会到综述应该是综述前人的研究，而不是自己构想一个研究。作为综述，应该是基于前人结果简要指出未来研究方向。如果中文摘要进行了修改，英文摘要作相应的修改。

回应：

非常感谢审稿专家对本文摘要撰写的细致指导。我们意识到，原摘要结构失衡的主要原因在于“研究展望(研究构想)”部分占据了过多篇幅。为此，我们在大幅精简这部分内容的基础上(详见“回应 1”及“回应 3”)，对中英文摘要进行了重写。

新版摘要参考了您建议的“背景/目的-方法-结果-结论”的结构，逻辑更清晰，重点更突出。按逻辑将句子组织为：

(1)背景：阐述忠实度的核心概念及其重要性。

(2)目的与方法：明确指出本文旨在填补的研究空白与主要工作，即通过系统梳理与比较分析展开论述。

(3)结果：提炼对现有测量方法及其改进策略的核心发现与评述。

(4)结论：基于前文分析凝练未来研究方向，并强调其对推动行业标准化实践的意义。

修订后的中文摘要内容为：“**心理治疗的忠实度(fidelity)**指心理治疗按预定方案实施的程度，是保障临床实践质量，验证疗效可靠性的关键指标。为填补国内学界在心理治疗过程质量评估中忠实度研究的空白，本文系统梳理了其概念构成与测量必要性，比较了常用测量方法的优势与局限，并评述了现有改进策略的贡献与不足。综述现有研究发现：观察性编码、

书面调查法和行为排练等方法虽各有特点，但存在耗时耗力、适用范围有限及评分一致性不足等问题；近年来，围绕测量框架建设、干预措施精确描述和计算机辅助评估的探索，在一定程度上提升了忠实度测量的规范性与效率，但仍难以兼顾长期存在的“高成本”与“评估不一致”问题。未来研究需进一步完善忠实度测量框架，建立统一知识库，并持续推进自动化评估工具(如大语言模型)在心理治疗情境中的实证检验，从而推动心理治疗过程质量评估从理论倡议走向标准化实践。”

相应的英文摘要为：“Fidelity in psychotherapy refers to the extent to which treatment is delivered as intended. It is an important indicator of clinical quality and of the reliability of therapeutic outcomes. To address the limited domestic research on fidelity in psychotherapy process evaluation, this article reviews its conceptual structure and the need for its assessment, compares common assessment methods, and discusses current strategies for improvement. Existing studies suggest that observational coding, written surveys, and behavioral rehearsals each have distinct features and limitations. Across current approaches, major challenges include high time and labor costs, limited applicability, and low scoring consistency. Recent efforts to develop assessment frameworks, specify intervention procedures, and introduce computer-assisted assessment have improved standardization and efficiency to some extent. However, the long-standing problems of high cost and inconsistent assessment remain unresolved. Future research may further refine the assessment framework, develop a more systematic knowledge base, and continue to explore and validate automated tools such as large language models in psychotherapy settings, thereby advancing the quality assessment of psychotherapy processes from theoretical advocacy to standardized practice.”

意见 3：【批注位置：正文“6 研究展望”首段】“并在此基础上试图提出一个整合性的测量方案，以供参考”【意见】建议删除。建议：研究展望：只保留 6.1 的三个探索方向，加最后一段话概括。供参考。

【批注位置：正文 6.2 节】“6.2 一个基于大语言模型技术的忠实度测量方案”【意见】建议删除

【批注位置：正文 6.3 节】“6.3 从忠实度测量到心理治疗质量评估体系的建立”【意见】建议删除

回应：

非常感谢审稿专家对本文“研究展望”部分内容的建议。经慎重考虑与修改，我们已在返修稿中删除了原“6.2 一个基于大语言模型技术的忠实度测量方案”及“6.3 从忠实度测量到心理治疗质量评估体系的建立”两节内容。

当前版本的研究展望仅保留了框架构建(6.1 持续完善忠实度测量框架)、标准制定(6.2 建立忠实度统一测量知识库)与技术应用(6.3 使用大语言模型技术辅助忠实度测量)三个层面的未来探索方向。此外，在结尾呈现一段总结性的概况论述，旨在强调忠实度研究在当前心理健康服务体系中的重要价值。

需要补充说明的是，为了兼顾第一轮审稿中编委 1 的初审意见以及您(审稿人 1-主要建议 4)关于“忠实度”与“循证理念”之间内在联系的关切，我们在返修稿的 6.1 节末尾，补充论证了从“循证实践(EBP)理念”视角理解忠实度测量框架的必要性。具体补充内容如下(返修稿 6.1 节末段)：“最后，心理治疗忠实度框架的有效解释和应用，有赖于循证实践(evidence-based practice, EBP)理念的指导。循证实践理念强调，需在充分考虑来访者的特征、价值观与偏好的前提下，将当前最佳的研究证据与治疗师的专业技能进行整合(American Psychological Association, 2021; APA Presidential Task Force on Evidence-Based Practice, 2006; Melchert et al.,

2024; 杨文登 等, 2017; 杨文登, 2012), 这与忠实度的评估逻辑一致, 二者均需以方案实施过程中的“证据”为核心。一方面, 忠实度测量是循证心理治疗得以成功实施、维持和扩展的前提(Bond & Drake, 2020; Marriott et al., 2025; Cook et al., 2017)。另一方面, “什么内容值得被忠实实施”本身需要由循证证据来界定。没有循证研究提供依据, 忠实度评估将失去参照, 容易沦为对治疗过程表面形式的记录。McLeod 等人(2013)进一步阐释了二者之间的关系: “依从性”评估了治疗师是否选择并实施了最佳的研究证据; “技术胜任力”考察了治疗师的临床专业技能; 而“基础胜任力”则可以反映出治疗师是否真正地结合了来访者的特征、价值观与偏好。由此可见, 在循证实践理念的指导下持续完善忠实度测量框架, 对于提升临床实践质量, 推动心理健康服务的持续改进具有重要意义(Goldman et al., 2001)。”

总体意见: 修改后发表。

回应:

非常感谢您的认可! 同时, 也非常感谢您在两轮审稿中给予的专业、细致的审阅与指导。得益于您的宝贵意见, 相信稿件的整体质量将得到显著提升。

审稿人 2 意见:

作者对论文进行了进一步补充和完善, 逻辑更加清晰。一个小问题: 图 1 似乎没有体现 4 个部分之间的关系, (4)中还有保真度的表述没有改过来。

回应:

非常感谢审稿专家对本文修改质量的肯定, 也感谢您耐心指出本文的细节问题!

首先, 对于原稿件中的疏忽我们深表歉意, 在原图 1 的(4)中确实遗漏了将“保真度”修改为“忠实度”的工作。在此次返修中, 我们已经对全文进行了再次的逐字排查, 确保不再出现类似的术语误用现象, 非常感谢您的指正。

其次, 针对您提出的“原图 1 中 4 个部分之间关系未充分体现”的问题, 由于考虑到综述的体例以及篇幅限制, 结合审稿专家的综合建议, 我们在最新的返修稿中已经删除了对“一个基于大语言模型技术的忠实度测量方案”的探讨内容(包含原图 1 及相应的小节内容)。

尽管该部分已被删除, 但为了更好地回应您的关切, 我们仍在此向您详细解释原构想中 4 个部分的逻辑联系, 以及我们最终决定删除该内容的考量。

原方案旨在针对忠实度测量长期存在的“低一致性”与“高成本”问题提供解决思路。方案流程包含框架设计、知识库建设、测量实施, 至平台建立四个阶段。文字(1)框架设计阶段、(2)知识库建设阶段、(3)测量实施阶段、(4)平台建立阶段与原图 1 的序号一一对应, 序号顺序为时间先后顺序, 在图中亦通过单向箭头标示。在内部关系层面, 前三个阶段表现为递进关系: 首先确立理论框架与方法学基础(框架设计阶段); 其次, 知识库承接框架, 并将其转化为具体的行为判定规则(知识库建设阶段); 最后, 在执行中进一步将知识库编译为大语言模型可执行的结构化指令, 完成从理论到实践的转化(测量实施阶段)。而第四阶段(平台建立阶段)不仅是前三个环节的集成, 更引入了反馈机制。通过收集实践、应用中的客观测量结果与多方(即用户)的主观反馈, 平台数据可用于持续修正初期的评估框架、扩充知识库内容体系, 并优化大语言模型的编译指令。总体而言, 我们希望这四个阶段能共同构成一个支持迭代优化的心理治疗忠实度评估体系。

然而, 在本轮修改过程中, 结合审稿专家的意见与我们的反思, 我们认为该测量方案虽然经过了前期的多方讨论, 但目前仍停留在构想与初步验证阶段, 尚未产出显著的实证数据支持。对于一篇以系统梳理前人研究为主的综述性论文, 过大篇幅地介绍属于作者构想的原创性工作, 不仅偏离了综述的主旨, 也可能冲淡文章本身的核心论点。将落脚点收敛于“未

来的探索方向”，可能更能启发该领域读者的思考。我们相信，即使删除了该方案细节，当前稿件依然保持了足够的信息量，且仍不违背我们将“忠实度”核心概念正式引入国内学界的初衷。

最后，再次感谢您的建议，您的两轮宝贵建议对本文质量的提升至关重要！

第三轮

编委 1 意见：

本稿围绕心理治疗忠实度测量中的核心瓶颈问题展开，选题具有一定前沿性与实践价值。经两轮修改后，作者已对综述结构、摘要写作及研究展望部分进行了实质性调整，基本符合本刊对问题导向综述论文的要求。外审意见已得到较为充分的回应。

目前稿件仍存在个别表述与综述定位不完全一致(如“构建方案/系统”的残留)及个别结构前后不一致的问题。建议作者在进一步收紧论文定位、统一全文表述后予以录用。

回应：

衷心感谢编委专家对本文的认可与耐心指正。对于前一轮返修的大幅调整后，正文中仍残留了少量与综述定位不符的表述，以及个别结构前后不够统一之处，我们深表歉意。

在此次修订中，我们组织了多人对稿件进行了反复通读，重点检查正文是否仍存在“构建方案”“建立系统”等残留表述，同时对各章节之间的专有名词、衔接与逻辑进行进一步统一。鉴于本次修改仅涉及部分表述或语序的微调，为方便后期审阅，本轮我们仍用第二轮的“黄色”对稿件的修改内容进行标注，并逐条说明如下。

具体改动包括：

(1) “1 引言”末尾段：**将原句**“.....最后在此基础上提出一个基于大语言模型技术的、整合性的忠实度测量方案设想”**修改为**“.....从**框架完善(6.1 节)**、**知识库建设(6.2 节)**与**大语言模型辅助测量(6.3 节)**三个方面，对未来研究方向进行展望。”修正后，引言部分在全文结构与写作目的上的概括，与稿件的整体定位保持一致。

(2) “6 研究展望”首段：**删除原句**“同时，将框架和技术真正落地，也离不开政府机构或行业组织的协同支持。”由于正文并未就“政府机构”或“行业组织”的作用展开论述(原论述已在第二轮返修中删除)，删除该句避免了本节首段引入内容与后文(三个展望方向)脱节，增强了文章前后结构的一致性。

(3) “6.2 建立忠实度统一测量知识库”末尾段：**将原句**“将这些共识标准整合为通用的评估工具或平台，是改进忠实度评估的关键。”**修改为**“**将这些共识标准进一步整合为通用的评估规范与知识资源，是改进忠实度评估的重要方向。**”本处修改旨在进一步规避“工具”“平台”等容易引发“方案建构”联想的表述，紧扣文章的综述定位。

(4) “6.3 使用大语言模型技术辅助忠实度测量”首段：**将原句**“高昂的人工评估成本也阻碍着整个评估体系的运转。因此，借助算法的力量将人工从忠实度测量中解放出来，已经成为了更具优势的选择。”**修改为**“**高昂的人工评估成本也限制了忠实度测量的常规运转。因此，借助算法减轻人工编码负担、提升大规模评估效率，逐渐成为值得重点探索的改进思路。**”本处修改主要是为了使论述更审慎、客观，确保技术应用部分的表述与全文的探讨语气保持一致。

(5) “6 研究展望”末尾段：**将原句**“.....在多方共同努力下，将有望建立起更完善、全面的心理治疗质量全流程评估体系，为我国心理健康服务领域的有序发展提供坚实支撑。”**修改为**“.....**汇聚多方力量持续完善忠实度测量体系，我国心理治疗过程质量评估的规范化水平有望得到进一步提升，进而保障心理健康服务领域的长远、有序发展。**”本处修改使全

文结尾与本文核心主题——“心理治疗忠实度测量”的对应关系更为紧密，避免结论过度延伸至本文主体论述范围之外的“全流程评估体系”。

再次感谢编委专家提出的宝贵意见，期待本轮修改能够妥善回应您的关切！

编委 2 意见：

这篇文章的选题是很重要的话题。作者已经做了全面细致的修改，可以刊发了。

回应：

十分感谢编委专家对本文的认可！

主编意见：稿件经过多位专家点审阅，作者进行了认真的修改，达到了发表水平，同意发表。