

《心理学报》审稿意见与作者回应

题目：中国人整体性思维特征的情境评估法及其应用

作者：柯晓晓；齐惠紫；梁家辉；金欣园；高婕；张明霞；汪亚珉

第一轮

审稿人 1 意见：

意见 1：

在成语使用上，与个人在汉语言教育训练的程度有关，使用成语的多寡和具体内容能否作为思维特征的测量指标，具体算法在构想效度上证据是否充分；

回应：

本研究使用的是被试报告的各种类型成语数量的比例而非多寡作为思维特征的测量指标。我们在文章结果部分添加了报告数与测量指标的相关分析及相应说明，被试的报告数量与整体性思维的三个指标（整体比例、会意比例、主观比例）相关均不显著。证明使用报告成语种类的比例能够较好的排除被试的成语水平对测量结果的影响。

表1 报告倾向与整体性思维指标的相关

	报告数	合理比例	整体比例	会意比例	主观比例
报告数					
合理比例	-.65**				
整体比例	.08	-.39**			
会意比例	.14	-.27**	.26*		
主观比例	.05	-.34**	.62**	.53**	

注：**为 $p < 0.01$ ，*为 $p < 0.05$

意见 2：

思维的整体性和辩证性是什么关系，具有相对性吗，还是具有交叉或互相包含性，在很多典型案例中，整体性和辩证性是同时存在的；

回应：

我们查阅文献认为，整体性和辩证性思维是中国式思维的两种典型特征，整体性主要体现在中国人对事物之间联系的认识中，而辩证性则主要体现在对矛盾的认识中（侯玉波，彭凯平，朱滢，2016）。两者是相互交叉的，共同包含于中国人的思维方式中，这些特征成语中也都有所体现（卢铁澎，2008）。实际上，在初次的成语的编码过程中也存在部分被同时编码为两种类型的成语。根据审稿意见，我们内部讨论后更换了文章的写作重点和成语的编码方式，不再同时测量成语体现的两种思维特征，选取整体比例、会意比例和主观比例三个维度的指标侧重于评估整体性思维。

意见 3：

以问卷形式测量创造力的效度是否能得到保障。

回应：

研究使用的《威廉斯创造力倾向量表》已经经过了世界范围内大量研究的支持，虽然仍然无法避免量表测量的固有缺陷，但其信效度表现良好，从冒险性、好奇心、想象力、挑战

性四个方面划分创造力倾向适合作为中国式思维和创造力倾向之间关系初步探索性研究的检验工具。这一工具也是目前创造力研究中使用最多的工具,包括在中国被试中开展的研究。在这意义上,当前还没有找到更合适的工具。

.....

审稿人 2 意见:

意见 1:

论文的定位。在我看来,本研究应定位于中国式思维特征的测量方法的开发,而不是中国式思维特征和创造力等构念之间的关系辨析。正如作者在“自检报告”的研究亮点部分指出,最突出的创新性贡献在于“运用 VR 技术,创设出了能测试个体中国式思维特征的虚拟现实实验场景”。我完全同意作者“囿于技术手段限制,已有中国人思维特征的研究多采用量表施策”的判断,探索新的测量方法非常有必要。如果能结合新技术,开发出内容上兼具文化特异性,而且方法上兼具生态性、直观性、自然性与完整性的测量方法,将大大促进中国式思维方式的研究深度和精度。基于这点考虑,我建议作者重新考虑,是否有必要将研究重新定位于测量方法的探讨。从我的角度看,我觉得是有必要的。

1.1. 第一,从目前的文献综述部分内容看,无论是从技术手段,还是从测试材料看,采用 VR 技术,基于汉语成语故事测量中国式思维特征,都是非常新颖的。将研究定位于此,所撰写的论文将根据影响力和贡献度。

1.2. 第二,是否符合信度和效度要求,是判断一种新颖测量方法的最基本要求。就目前看,作者并没有报告这个测量的信度和效度指标(如区分度,聚合效度,预测效度等效度指标)。只有满足信度和效度要求之后,才能广泛应用。作者在讨论部分认为:“研究所使用的成语谜语测试系统在初步测试中表现良好,这为以后基于此研究开发形成相应的思维特征测量工具提供了初步数据基础”。但是,我并没有从论文中找到能反映信效度的数据指标。

1.3. 尤其关键的是,和既有的量表测量相比,这种新的测量方法在哪些方面存在优势,其独特性在哪里?

回应:

1.1.感谢您的建议,我们对论文定位及内容进行了重新调整,侧重于中国人整体性思维的情景评估方法的发展及初步应用。具体的修改内容如下:首先,我们进一步聚焦了评估的重点,不再测量辩证性思维,同时将整体性思维指标进一步细分为整体比例、会意比例和主观比例。其次,我们增加了对这一方法的信效度报告内容。最后,我们将于创造力倾向的关系研究作为本方法的初步应用进行了探讨。

1.2. 已添加信效度分析。

1.3. 情境任务法相比于量表法有优势主要体现在以下三点:首先,情境任务法使用虚拟现实技术来呈现仿真情境,以让被试在与情境的交互中完成测试任务。相比量表法而言,这显著提高了测验的生态效度。成语故事是对现实情境的抽象与概括,并且经历长时间的文化塑造最终稳定下来,这种抽象概括的情境与虚拟现实环境的设计特征(想象性,沉浸感及交互性)十分契合。因此用虚拟现实技术来呈现成语故事的生态学效度很好。其次,成语故事是研究中国人思维特征的良好材料。作为中国人的思维特征在本质上是形成于中国文化环境,也主要体现于中国文化环境。量表,尤其是自陈式问卷,在展示传统的文化环境上显然比虚拟现实环境要差。成语在某种程度上可以算是中国传统文化的精髓所在,是在文化环境中总结出来的活的思维展示,因此在“仿真”的文化环境中再现成语故事情景来测量思维特征,所选行为样本的代表性很好。最后,从测验技术上看,完全沉浸式的虚拟现实环境在诱发被试的思维活动时更为自然,在很大程度上避免了量表法可能带来的受测者的自我掩饰等偏差。

意见 2:

如果作者仍然希望将本研究定位于变量间关系探讨,如思维的整体性或辩证性特征和个体创造力之间的关系,我认为在研究模型以及数据层面都需要较大调整和补充。

2.1. 研究模型方面,我认为重点应该聚焦在中国式思维和创造力的关系上,即需要回答一系列问题:用这种新颖方式测量的中国式思维特征,在控制了其他重要因素的影响效应之后,是否显著影响创造力(主效应)?是如何影响创造力的(即中介或内在作用机制)?其影响是否会受其他因素所影响(即边界条件)?此外,正如作者所指出,“以往的大多数研究都暗示了辩证性思维方式与创造力之间的紧密联系”。那么,就两者关系探讨这个主题而言,本研究有何独特贡献?

2.2. 大五人格特征在本研究中的作用。在“1.5 研究目的和假设”部分,作者指出研究的目的之一是“人格特征独立于思维特征对创造力产生了影响”。如果本研究的主要目的是揭示中国式思维和创造力的关系,我觉得比较合理的研究设计是将大五人格作为控制变量,目的是展示控制人格特征的效应之后,思维特征的效应仍然显著,而不是将其纳入作为主要研究内容之一。否则,就是偏离研究主题,事实上我本人就感到有些困惑。

2.3. 第三,就中国式思维和创造力关系探讨这一主题,我觉得仅依靠一个研究,数据显得不足。尤其是,正如作者所指出,威廉斯创造力倾向量表测量了个体的创造力倾向,这是相对稳定的个人特质,而不是个体的创造力表现。如今停留在这个层面,实际上只是探讨了两个相对稳定的个体差异(中国式思维特征和创造力倾向)之间的关系,而无法全面回答中国式思维和创造力之间究竟存在哪些关系。要达到这一目的,则需要测量个体的创造力表现,比如通过评价个体的创造产品,来衡量其新颖性,灵活性,实用性,完整性等常见的创造力指标。

回应:

我们经过分析采纳了审稿人有关定位上的意见,不再将研究定位于变量间关系的探讨。仍保留和与创造力倾向的相关分析,并进一步根据被试整体性思维特征分类进行了创造力倾向中好奇心得分的 T 检验,其目的是验证整体性思维评估方法应用的可行性并初步探索整体性思维与创造力倾向之间可能的关系。

2.1. 我们不再重点讨论中国式思维与创造力的关系,仅从实证的角度为整体性思维对创造力倾向可能存在抑制作用提供论据。

2.2. 我们将研究重点重新定位于发展整体性思维情境评估方法后,将大五人格特征测试作为校标进行了讨论,探讨了整体性思维特征与人格方面的相关,以此作为本测验工具区分效度和聚合效度的部分指标。

2.3. 目前我们只是从已有的假设出发进行了初步的实证检验,验证了本评估方法的可行性和初步确定整体性思维特征可能对创造力具有抑制作用。对于中国式思维与创造力关系的研究工作还有待进一步深入,丰富创造力指标等问题,我们希望在后续的研究中进一步补充。

意见 3:

关于本研究的中国式思维的测量方法,我还有如下疑问。

3.1. 成语谜语库如何应用于实验被试的中国式思维测量?核心成语谜语库和自由创意成语谜语库之间,是什么关系?

3.2. 成语谜语测试是在什么阶段进行测试?

3.3. 成语编码:负责则编码的专家指的是谁?为什么可以认为是专家?专家根据什么标准编码?编码一致性有多高?

3.4. 作者指出,“在辩证型成语数量和整体型成语数量中,不包含那些既被标注辩证型,

也被标注为整体型的成语”。那么，这些同时被编码为辩证性和整体性的成语，是如何处理的？为何这么处理？同时被编码为辩证性和整体性的成语所占比例，是否有独特含义，比如可能反映了个体两种思维方式的平衡性程度？

回应：

3.1. 我们在文章中进行了补充说明，核心成语库是预先从网上的分类成语库中选取、进行形意转换创意，并按照静态组合，动态展示两类进行平衡的成语，收录的是研究者有意引起被试联想的成语谜语。自由创意成语库则是记录被试在实验过程中根据虚拟情境的场景内容自由联想到的除核心成语库之外的合理的成语。通过这种方式，保证被试报告的所有符合情境信息的成语都加入成语谜语库，在后续的编码中两类成语标准一致，将被试汇报的成语对应成语库中的编码计算各类成语比例作为整体性思维的指标。

3.2. 成语谜语测试可以针对基础成语有所了解的人群在任何阶段进行测试。

3.3. 我们重新细化了整体性思维特征的指标，选择了三个维度对成语进行重新编码，编码的方式较为简单，根据成语词典含义进行划分，不涉及高深的语言知识。本次编码由一名心理学研究生单独负责。

3.4. 在初次编码中，同时被编码为辩证型和整体型的成语约占 7.19%，此部分成语计入被试报告的成语总数，但不计入被试的辩证型和整体型成语数量计算比例。两种成语存在重叠实际上符合辩证性与整体性思维可能同时存在这一现象。在重新编码时没有对辩证型成语进行编码，仅侧重于整体型成语，是因为本研究目的主要是发展整体性思维的评估方法，成语中蕴含的辩证性思维特征的测量等在后续的研究中可以进一步尝试。

意见 4：

控制变量。成语测试的成绩可能受到被试的汉语水平（特别是成语回忆能力）的影响，是否需要将此纳入作为控制变量？毕竟被试的成语测试成绩差别很大（如作者指出：大多数被试在系统中都能够说出约 30 个成语，最多的能说出 133 个）。此外，临场感和研究所涉及的主要变量之间是否存在显著相关？是否也应该作为控制变量？

回应：

考虑到成语的使用可能会受到个人文化水平等因素影响，本研究使用的是被试报告的是不同类型成语报告的比例而非数量作为思维特征的测量指标，对于被试汉语水平的依赖较小。我们在文章中也添加了报告数量与所选指标的相关分析，之间不存在显著相关。

另外测量临场感的目的是对于研究设计使用的虚拟现实环境质量进行评估，避免不良测验环境对被试产生干扰，得分证明本研究所使用的虚拟现实环境质量良好，能够使被试产生较好的沉浸感。临场感得分与研究所涉及的主要变量之间相关均不显著。我们在文章中也进行了补充说明。

意见 5：

表 3 中显示，开放性与辩证性的相关系数 $r=0.25$ ，是否显著？

回应：

指标修改后，这部分数据已删除。

意见 6：

这种基于中文成语的测试，对汉语能力有很大依赖性，如何用于跨文化研究？

回应：

目前已有许多基于西方文化环境发展并已应用于跨文化研究的思维方式测试，但这些研究中文化情境差异的体现较为不足。本研究是在前人研究中已证实的中国式思维与西方人传

统形式逻辑思维不同的基础上,初步探索在传统文化情境中使用特殊文化符号测量中国式思维特征的可能性。进一步的研究方向可能是比较传统文化情境和非传统文化情境下测量到思维特征的是否有差异,从而进一步理解文化对于思维的影响,而非直接将中文测验应用于跨文化研究。

第二轮

审稿人 1 意见:

意见 1:

该项研究也是一种新的尝试,具有一定的理论与创新价值。修改之后,论文质量有了一定的提升。但也还存在一些问题需要思考或值得注意:关于整体思维、辩证思维、创新思维三个概念之间的关系,是一个关于思维结构的重要问题,需要进行深入地思维探讨。就本文来说,从文献综述、理论分析、数据分析,均不足以解决这一理论问题,所以后续数据分析中的处理方式,理论构想效度算不算高;

回应:

已有许多研究者基于实证及经验对三者的关系提出了思考,有研究者认为辩证性思维在发现问题的过程中具有重要作用可能会对创造思维具有促进作用(Arlin, 1976; Arlin & Levitt, 1998; Benack, Basseches, & Swan, 1989; Rothenberg, 1996),而中国人场依存认知风格可能是由于难以将目标从整体中区分出来的整体性思维结果(Ji, Peng, & Nisbett, 2000)及大多研究在场依存与创造力的关系上则展现了负相关的结果(武欣,张厚粲,1997;李寿欣,李涛,2000)也暗示了整体性思维可能与创造思维具有某种关系,本研究的假设也是基于这些结论推导而成。

但从前人研究成果来看,目前三者的具体关系仍缺乏权威有效的理论结论,这种思维结构的探讨一定程度上也受限于目前中国人思维特征测量的局限性。作为一项发展新方法的探索性研究,本研究也无法直接解决这一理论难题,但希望通过评估方法探索为后续的理论研究提供新的可能,并根据进一步的理论发展不断进行调整完善。

以上内容我们在讨论部分进行了补充说明。

意见 2:

论文中谈到计算成语使用的总数与三个比例的相关均不高,这一结果能否说明三个比例排除了成语使用数量的影响呢?不一定,因为被试使用的成语总数分散性很大,有的个位数,有的三位数,被试击中成语总数的标准差很大,这无疑会影响到相关系数的计算(或许,应该排除成语击中总数太少的被试,再进行技术会好一些);

回应:

感谢您的建议,我们在处理数据时没有删除报告数较少的被试,主要是考虑报告数量除了受到被试成语知识的影响也可能体现了部分不同的思维特征,按照报告数删减被试可能会造成代表性降低或结果偏差。为了消除分散性造成的影响,我们尝试按照两种方式进行了数据剔除:一、排除报告总数为个位数的被试3人,结果见表1;二、排除报告数 ± 2 个标准差以外的被试5人,结果见表2。两种处理方式下,成语报告数量的影响也均不显著。

表1 报告倾向与整体性思维指标的相关 (N=87)

	报告数	合理比例	整体比例	会意比例	主观比例
报告数					
合理比例	-.66**				
总体比例	.13	-.38**			
会意比例	.12	-.24**	.17		
主观比例	.07	-.35**	.57**	.48**	

注: **为 $p < 0.01$, *为 $p < 0.05$

表2 报告倾向与整体性思维指标的相关 (N=85)

	报告数	合理比例	整体比例	会意比例	主观比例
报告数					
合理比例	-.53**				
总体比例	-.02	-.37**			
会意比例	.19	-.29**	.25*		
主观比例	.05	-.37**	.62**	.52**	

注: **为 $p < 0.01$, *为 $p < 0.05$

意见 3:

在写作中,“结果”部分明显的夹叙夹议,未能保持结果呈现的客观性。

回应:

已在文章相应位置进行修改。

意见 4:

关于创造力倾向的问卷测评,使用的人多不代表该量表效度就高,该领域中有效工具本来就少,不排除很多研究者在无其他工具可用情况下,使用该工具解决燃眉之急,创造力或创造力倾向的测评,迄今并未得到有效解决。相比,托兰斯的工具效度更高,但使用起来更为繁琐,所以被采用的频数不高。

回应:

再次感谢您的建议,目前本研究只是对整体性思维特征与创造力可能存在的关系进行了初步验证。正如您所言,创造力测量仍处于探索阶段,有效工具较少。一方面来说为了进一步探讨整体性思维与创造思维的关系,在后续研究中,我们需要丰富创造力的指标;另一方面来说,本研究也希望通过特定文化情境中思维评估方式的探索为发展更有效的中国人创造力测验提供可能。对此已在文末进行了补充讨论。

.....

审稿人 2 意见:

意见 1:

可以看到,在修改稿中,作者只聚焦于整体性思维,而舍弃了辩证性思维的探讨。然而,在回复意见和正文中,我都没有看到作者对这样选择的理由做出解释。将文章重点调整为检验测量方法的信度和效度,并不意味着要舍弃测量指标本身。当然,我并不是说,这种测量方法一定要全面测量中国式思维方式(整体性和辩证性),只是希望看到作者为何这样考虑。

回应：

考虑到整体性主要体现在中国人对事物之间联系的认识中，而辩证性则主要体现在对矛盾的认识中（侯玉波，彭凯平，朱滢，2016）。当前使用的虚拟现实园林场景中，典型的矛盾冲突的信息物较少，相对于其他辩证性思维研究（如 Peng 等人在一系列社会认知研究中设置的冲突场景）对辩证性思维的诱发相对较弱。实际编码中辩证型成语的数量也较少，直接使用这部分数据进行中国人思维解释，可能会造成中国人辩证性思维较弱的误区。此外如正文“1.2 成语故事中的整体性思维特征”部分的补充说明，辩证性与整体性思维的结构关系涉及更深入的思维特征模型，本研究不将其作为讨论重点。因此本研究聚焦于整体性思维，对于辩证性思维的测量有待进一步探索新的情境设置及理论深入再加以补充。

意见 2：

作者提出，整体性思维特征包括整体性、会意性和主观性三个指标。首先，我非常赞赏作者能明确提出自己的观点。但我在理论和实证依据两个方面，有一些疑虑。

2.1.理论建构层面。我注意到，作者在论证过程中并没有引用任何文献（正文“1.2 成语故事中的整体性思维特征”）。因而我的疑问是：（1）在已有文献中，研究者是如何界定整体性思维特征的？将整体性思维界定为整体性、会意性和主观性三个特征，是否有依据？（2）这三种特征，是否只是基于词汇的思维方式测量才能测量出来，还是用其他手段（比如量表法）也可以测量？（3）这三个特征之间存在什么关系？三种特征中，是否存在哪种特征（比如整体性）是整体性思维的核心特征？整体性思维一定会体现出会意和主观特征吗？三个指标能否聚合为单分数（或者使用潜变量）？

2.2.在实证层面，如何理解和解释三个指标之间的关系。从表 4 可以看到，三个指标存在显著正相关，然而考虑和其他变量的相关时（如表 5 和表 6），会意性和主观性和大五人格各维度、创造力倾向各维度均无显著相关，是否说明这两者并非整体性思维的特征？另外，由于没有报告三指标和场依存型-独立性的相关矩阵，无法判断会意性和主观性的相关情况。

回应：

2.1.（1）已在正文“1.2 成语故事中的整体性思维特征”部分添加了补充说明。为了避免整体性指标与整体性思维概念混淆，我们经过探讨后决定将成语思维中体现整体环境内容和被试总体心境体验的“整体性特征”更名为“总体性特征”。本研究所提出的总体性、会意性及主观性三个特征，是基于 Choi（2007）以及侯玉波，彭凯平，朱滢(2016)分别提出的两种思维结构，删除辩证性特征的部分因素后，结合成语内容进行分析整理而成。相对于原结构特征，因素之间的概念重叠较小，在成语编码时的操作性更强。

（2）本研究依据成语内容对这三种特征进行了具体解释，使其更能代表文化特性，但其本质并不局限于词汇思维，总体性主要代表“由点及面”的思维特征，会意性主要指“由表及里”或“由此及彼”的思维特征，而主观性则是在思维过程中将个人经验体验融入的程度。这些特征通过量表等方式也许能够测量，但目前缺乏有针对性的工具。

（3）三种特征是整体性思维的不同维度，从目前的研究结果来说，不能断定是否存在哪种特征是整体性思维的关键特征，整体性思维并不一定同时体现出全部三种特征，各种特征之间的具体关系有待进一步的研究进行补充。

2.2. 与大五人格的相关来看，主观比例与神经质负相关显著，会意比例与宜人性的负相关显著，具体内容我们在正文“3.3.2 大五人格”部分进行了讨论，这两项结果均符合我们的假设，一定程度上证明了整体性思维特征测量的聚合效度。会意性与主观性与创造力倾向各维度均无显著相关，可能是由于本研究选用的测量工具仅测量创造力的倾向而没有测量创造力的表现，注重事物内在联系和注重主观体验更可能影响创造过程及结果而非创造倾向，我们在结论部分对其进行了补充讨论。已在文章相应部分补充了整体性思维指标与镶嵌图形分

数的相关结果，场独立-场依存的认知方式仅与总体比例显著相关，场依存认知易受背景环境影响与总体比例反映的注重对象与环境联系的关注点是相符合的，而与是否关注事物内在含义和主观体验则没有必然联系。表明了三个指标具有一定的区分度，测量了不同的思维特征。

意见 3:

成语的编码。

3.1. 一般来说，编码质量是很关键指标。只依赖单一编码者对被试报告的成语进行编码，结果是否可信和可靠？毕竟一个成语究竟属于哪种类型，有较大主观性。

3.2. 根据什么来判断被试所报告的成语是否合理？另外，既然存在合理与否，是不是意味着基于词汇的思维方式测评，存在事先规定的正确答案？如果被试报告了“不合理”的词语，是否说明该个体打破既有规则的倾向性越强，也是有意义的？

3.3. 整体性、会意性和主观性三个指标的衡量，是以合理成语数作为标准，为什么？如果以报告成语数作为标准，结果如何？

回应:

3.1. 本研究没有进行多人编码的原因是由于使用的虚拟现实场景进行了部分成语的形意转换，编码需要熟悉虚拟现实场景，如“按兵不动”这一成语，仅按照词典意义划分属于会意词语，但在本场景中将其创意转换后形象表达为“一只手将士兵按住不动”，需要编码为会形成语。我们重新细化后的编码操作性定义较为精确，编码的方式简单，仅需要根据场景内容和成语词典含义划分描述细节/整体，是/否具有引申含义，是/否包含主观体验，不涉及高深的语言知识，努力降低单一编码者的主观影响。

3.2. 任何与虚拟现实场景中涉及的客体形象、环境氛围、被试的行为表现相符合的成语都被判定为合理成语。如描述场景中具有楼、台、山、河、花、草等形象，概括环境的整体氛围，表达被试游览时心情感受、行为动作的成语等。仅将环境中没有可以引起联想的线索、内容与环境无关的成语，如“贼眉鼠眼”“良药苦口”（园林环境中没有人物、老鼠、药等形象线索）等，作为不合理成语。这种方式并不意味着事先规定了正确答案，相反我们创立自由创意成语库收录被试报告中事先没有预设的成语正是为了避免正确答案的限制。正如“正文 3.4 测试的应用检验”中所报告的，合理比例与创造力倾向中的好奇心是显著正相关的，合理比例代表着报告成语与当前环境体验符合的程度，一定程度上可能反映了被试的探索欲。

3.3. 以合理成语数为标准的目的是为了减少被试追求报告数量进行机械背诵造成的偏差，数据处理时更关注的是可能由环境诱发的成语，这些成语更能反映被试在所塑造出的文化情境中的思维，如果以报告成语作为标准则较难排除背诵带来的影响。

意见 4:

测量信度。作者将“成语库中的收录的成语作为项目，按被试是否击中进行计分”，将此定义为测量的内部一致性 α 系数。首先，我不理解 α 系数是如何计算出来的。另外，击中与否是不是正确率指标？是不是合理成语的判定？如果是，为什么正确率/合理性可以作为信度（内部一致性）的指标？

回应:

已在文章相应部分进行补充说明。本研究中内部一致性系数代表被试能否根据 38 类典型情境报告出合理成语的一致性，结果表明所使用的各类情境能够较为一致的诱发被试的合理成语联想。根据当前情境信息“言之有理”的报告结果均被判定为击中，对报告成语的合理性进行限制是为了聚焦由情境信息诱发的成语中所体现的文化思维。

意见 5:

结果报告。建议在统一的一个描述统计表中报告研究所涉及到的所有变量之间的相关，从而提供更丰富的信息。比如，表 6 和表 7 都报告了“合理比例”这个指标，但表 5 中缺失。成语报告数和其他指标的相关也有意义，但只在表 4 中出现，而其他表格却没有。

回应:

感谢您的建议，将所有变量相关放入一个统计表中表格过大，已分别在表 5 和表 6 中添加完整。

第三轮

审稿人 1 意见:

论文采用虚拟场景刺激条件下成语使用的选择，作为评估中国人整体性思维方式的测评方法，是一种有创新性的探索工作，具有一定的意义。作者结合评审专家的意见，对文稿进行了耐心修改，文稿质量有了较大提升。但是，由于其研究本身的构念、数据收集方法和数据处理方法是研究一开始就确定的，并不能在修改中重新改变，所以 2 位评审专家一开始提出来的有些问题并不能在修改中就能解决，最关键还是该种方法的信效度问题，尤其是效度问题。作者就其自己的研究设想进行了阐释，在一定程度上表达了作者本身的研究思维和逻辑。作为探索性问题，提供给读者去思考和讨论也是有意义的事情。

回应:

感谢您的建议，为了对本研究的信度进行进一步的补充，我们增加了两名成语编码者，计算了评分者信度，结果补充在了正文“3.1 描述统计与信度检验”中。作为一项探索性研究，确实存在理论建构和数据支撑等方面的不充分，希望能够在进一步的研究和使用中进行检验和补充。

审稿人 2 意见:

意见1:

理论上，我仍然不能理解，本研究所提出的三维度结构（总体性、会意性和主观性）和前人的维度划分（AHS 量表，中国人整体思维方式量表）的区别与联系。在我看来，总体性这个维度是核心，和前人两个维度结构（如关注点，联系性）也是相符的，会意性和主观性并不是整体思维方式的核心，而是因为采用虚拟现实的成语联想这种测量方法所特有的。从研究结果也可以看到，总体性和其他变量存在较为稳定的相关关系，后两个指标则无显著相关或相关关系不稳定。因而，我认为作者对这种测量方法的本质，界定不清晰，理论建构上尚不清楚。

回应:

从理论基础的角度来说，会意性和主观性与前任整体性思维划分维度的关联确实不如总体性清晰直观。但从实践的角度而言，我们在对被试报告的成语进行编码时发现仅从成语体现的是整体环境信息或总体心境这一个标准可能并不足以概括成语表达中体现的整体性思维，部分成语虽然描述的是具体的事物特征，但也反映了被试对具体事物之间、表象与内涵之间、以及“个体”与环境之间的交互观念，我们在正文“2.6 数据处理”中添加了具体的例子来帮助说明选取编码标准的思路。在对这些成语思维进行分析后，认为其符合前人研究中对

“联系性”和“因果观”的定义，因此选择将这两项特征纳入整体性思维评估。

正如专家所指出的，本研究的结果中总体性指标与其他变量的相关比会意性和主观性更稳定，表明总体性可能是成语整体性思维评估的核心指标，但从这一结论出发排除会意性和主观性也可能造成对整体性思维的误判，并且依据成语的实例分析，相对于总体性更多的体现整体性思维“关注点”这一表面特征，会意性与主观性更可能反应“联系性”和“因果观”的深层次特征，因此在本研究中予以保留。这些指标间的关系还有待在后续的理论 and 实践探索中辨明，在讨论部分我们也对这部分内容进行了补充说明。

意见 2:

方法上。

2.1. 开发测量方法时，测量的信度是最重要的指标之一。对于本研究来说，真正的测量指标是成语联想的总体性、会意性和主观性，应该报告得到这三个测量值时，所依赖的测量手段是否可信/稳定。“根据所使用的虚拟现实场景中涉及的 38 类典型文化要素进行情境划分，按照被试能否根据该情境报告出合理成语进行 0-1 计分”所得到的测量分数，实际上是反映成语报告“合理数”这个指标是否可行，和报告成语的总体性/会意性/主观性衡量指标，没有必然联系。另外，就测量方法开发而言，小规模的单一样本，提供的信息量也是不足的。

2.2. 报告成语总数及其性质（总体性/会意性/主观性），是否测量了该被试的整体性思维特征，而不是基于虚拟现实的词语联想测量方式所测量到的整体性思维特征？如果从这个角度考虑，这个指标也是有意义的。

回应:

2.1. 感谢您的建议，我们必须承认由于前期的研究方向并不在于评估方法的确定，我们在信效度问题上考虑的不够充分。为了更好的弥补这一问题，我们增加了两名编码者，对整体性思维指标的评分者信度进行了分析，结果补充在了正文“3.1 描述统计与信度检验”中。

2.2. 采用虚拟现实技术构建这样一个典型的中国传统文化园林的情境，目的就是要提高中国人思维方式研究的生态学效度。我们设想在这样的环境中来观察被试思维的整体性，会意性与主观性会更加自然，具有情境性，生态学效度也好些。虚拟现实环境的内容，也就是传统中国文化园林的构建是被试后续心理加工的基本背景，从虚拟现实环境的沉浸感体验上来看，这种园林文化情境的体验还是很强的。根据以上基础设想，本研究更关注的是在利用虚拟现实构建的文化情境下中国人展现的思维特征，而非跨情境下对整体性思维的认同程度，这也是与前人量表测量方式的主要区别之一。我们认为使用合理成语作为指标能够更好的避免被试在场景中进行机械背诵带来的偏差，更聚焦于特殊文化情境下的思维特征。依据编码标准，只有报告的成语既不涉及环境中的任何物体、不符合情境的延伸含义、也不体现被试自身当前的行为表现时才会被排除，例如“盘古开天”。这些成语相比于受环境刺激诱发，更可能是被试的背诵结果。当然不可否认这部分成语可能反映了思维特征偏向在学习和记忆等方面的差异，在后续的研究中对不同环境及不同条件下的成语报告思维特征进行对比也是有价值的研究方向，但并非本研究的关注重点。

.....

审稿人 3 意见:

意见 1:

“本研究将成语中蕴含的整体性思维特征概括为总体性、会意性和主观性三个指标”——审稿人认为这三个指标中的“总体性”可以接受，为什么“会意性”和“主观性”可以体现整体性思维特征呢？根据表 1 和表 2（此处作者编号错误）的内容，读者可以认为整体性思维特征

与会意和主观这两个指标没有多大的联系,会意强调的是对语义的理解凭什么就一定要基于整体性思维呢?即使是呈现虚拟的场景图片,同样有可能对某些细节进行顿悟后领会场景的语义。如果作者坚持其观点,至少还应该进行更令人信服的阐述。

回应:

感谢您的建议,我们在正文“2.6 数据处理”中增加了这三个指标的具体选取过程及相应的成语用例,以便更好的说明指标的选取思路。根据前人研究,整体性思维指倾向于用联系的、整体的观点看待问题,倾向于强调事物之间的关系,往往会将问题与所处的背景相结合(侯玉波,朱滢,2002)。在本研究中总体性的编码标准主要体现的是以概括的视角看待事物及环境、将目标与背景看作是一个整体,相对来说这一标准是最为简明直观的,但在实际编码过程中,我们认为仅用这一划分难以涵盖所有体现整体性思维的成语,部分成语虽然描述的是事物的细节线索,但却反映了表象细节与事物内涵或背景环境的联系以及个体经验感受与背景环境交互作用。如“点睛之笔”、“一叶障目”、“刀枪不入”等,这两类成语也分别体现了“表象”与“内涵”的联系、“人”与“环境”的联系,符合前人研究中整体性思维定义,因此将其纳入整体性思维评估指标,目的是希望能够更广泛地涵盖成语可能体现的整体性思维特征。这种选取的方式实用性更高,但理论严谨性较弱,我们仍希望保留这两个指标作为后续研究的参考,并在后续研究中进一步对其进行厘清和优化,我们在讨论部分增加了对这部分局限性和未来展望的具体讨论。

意见 2:

“侯玉波、朱滢和彭凯平(2016)编制的中国人整体思维方式量表”——既然该研究参考了中国人整体思维方式量表,为什么不考虑将这一量表作为该研究的校标,进一步检验效度呢?或者再利用解决其他思维问题的表现考察效标效度呢?然而,作者将镶嵌图形测试分数作为校标这一做法可能有一定问题,镶嵌图形测试主要是从视知觉层面反映认知风格,这与思维方式恐怕还是有差异的。另外,为什么不将联系性、变化性和矛盾性三个因素做相关分析呢?

回应:

中国人整体思维方式量表在编制时主要考虑尽量避免受情境的影响,题目主要涉及对自己的认知观念、对他人的认知观念、对社会事物和人际交往的认知观念(侯玉波,朱滢,彭凯平,2016)。从这一角度来说,该量表主要测试的是对于“整体性思维观念的认同”,而本研究则侧重于从具体的“情境感知与思维特征的关系”的角度来探讨文化情境下中国人实际思维特征。相比而言,镶嵌图形体现的场独立与场依存认知风格的划分在反应论的意义上会更符合些。在研究设计中具体考虑的理由有:1)都是基于特定的视觉环境刺激来测量人的思维加工方式的;2)都是在一个模拟仿真的空间环境中(场独立与场依存划分最早源于模拟飞行训练测试)来测量的;3)都是基于图形意象来评估人思维特征。

正如您所言,镶嵌图形测试反映的认知风格与思维方式之间仍可能存在一定的差异,但考虑到以上因素,我们认为以镶嵌图形测验作为本整体性思维评估方法的校标是较为合理的。以往关于中西方思维方式的差异研究也主要是从认知取向和归因模式的角度进行切入,研究对于事物之间联系的思维倾向(Morris & Peng, 1994; Ji, Peng, & Nisbett, 2000; Nisbett, 2004),您的意见也给了我们进一步的启发,在后续的研究中或许可以参照归因差异研究进行整体性思维方式的检验。据此,我们对讨论部分进行了补充。

我们在正文“1.2 成语故事中的整体性思维特征”中对于以往研究中提出的整体性思维相关维度进行了解释说明,前人所提出的思维结构是包含了辩证性思维的广义“整体思维特征”。而本研究的重点并非辩证性与整体性思维的复杂结构模型,因此仅选择体现“联系地、整体地看待事物”这一整体性思维关键特征的“关注点”“联系性”与“因果观”三个维度进行

操作化。没有对辩证性思维特征中的“变化性”和“矛盾性”进行编码的原因是考虑到整体性主要体现在中国人对事物之间联系的认识中，而辩证性则主要体现在对矛盾的认识中(侯玉波，彭凯平，朱滢，2016)。当前使用的虚拟现实园林场景中矛盾冲突信息物较缺乏，因此对辩证性思维的诱发相对较弱。尝试进行的实际编码中辩证型成语的数量较少，直接使用这部分数据进行中国人思维解释，可能会造成中国人辩证性思维较弱的误区。因此本研究聚焦于整体性思维，对于辩证性思维的测量有待进一步探索新的情境设置及理论深入再加以补充。