

基于 COSMIN 指南的创造力自评量表测量学属性综述性评价

A Systematic Review of the Measurement Properties of Self-Report Creativity Scales Based on the COSMIN Guidelines

王敏嘉¹, 高淑娟², 万海鹏^{3*}

¹³ 首都师范大学教育学院

² 首都师范大学教师教育学院

* dnvhp@163.com

【摘要】 本研究基于 COSMIN 指南, 对 9 个创造力自评量表的内容效度、结构效度、内部一致性及构念效度假设检验进行综述性评价。结果显示各量表内部一致性良好, 但内容效度均因未进行目标人群可理解性检验而评为“不充分”; 采用验证性因子分析的量表结构效度更优; 仅少数量表的构念效度假设检验达“非常好”标准。本文还结合生成式人工智能(GenAI)辅助测评工具的相关进展, 以期对创造力测评工具的选用提供参考。

【关键词】 创造力; 自评量表; COSMIN 指南; 测量学属性; GenAI 辅助测评

Abstract: This study evaluated 9 self-report creativity scales using the COSMIN framework. All showed good internal consistency but inadequate content validity. Scales using confirmatory factor analysis demonstrated better structural validity, with only a few achieving very good construct validity. The paper also discusses generative AI-assisted assessment, aiming to inform scale selection in creativity research.

Keywords: creativity, self-report scales, COSMIN guidelines, measurement attributes, GenAI-assisted assessment

1. 引言

创造力自评量表因施测效率高、适用场景广, 是创造力研究的主流测评工具。然而, 当前研究在量表选用上存在过度依赖单一信度指标、对内容开发与结构稳健性关注不足等问题, 而 COSMIN 指南为多维度量表评价提供了规范标准。结合 GenAI 重塑测评范式的趋势, 本文基于 COSMIN 指南对 9 个创造力自评量表进行综述评价, 并结合 GenAI 辅助测评展开讨论。

2. 方法

检索中国知网、万方及 Web of Science 数据库, 中文检索词包括创造力、创造性行为、自评量表等, 英文检索词包括 creativity、self-report、scale 等; 纳入聚焦创造力相关构念、以自评量表开发/修订/翻译为主要目的且报告测量属性的文献, 排除无法获取全文及非中英文文献; 采用 COSMIN 偏倚风险评估清单, 从内容效度、结构效度、内部一致性、构念效度假设检验四方面评价方法学质量及测量属性质量, 同时用 GRADE 法进行证据质量分级。

3. 结果

3.1. 纳入量表的基本特征与方法学质量 纳入量表的基本特征与方法学质量评价见表 1。

基金项目: 2025.1-2028.12, 2024 年度国家自然科学基金面上项目“面向社会性科学议题的在线合作论证学习评价模型与反馈支架构建研究”(62477034)

表 1 纳入量表的基本特征与方法学质量

量表名称	文献来源	类型	研究对象	题项	因子数量	因子描述	Cronbach's α	结构效度指标	COSMIN 方法学质量评价			
									CV	SV	IC	HV
大学生创造力量表	何少颖等人, 2015	自评	大学生	共 16 个题项; 采用李克特 5 点量表	3	(F1)发散性思维因子、(F2)才智应用能力因子、(F3)人格特征因子	0. 867	CFA+EFA	I	VG	VG	I
CMS(创造力心态量表)	Karwowski, 2014	自评	大学生、成人	共 10 个题项; 采用李克特 5 点量表	2	(F1)增长型心态、(F2)固定型心态	0. 65-0. 79	CFA+EFA; IRT	I	VG	VG	D
K-DOCS (Kaufman 领域创造力量表)	Kaufman, 2012	自评	大学生	共 50 个题项; 采用李克特 5 点量表	5	(F1)日常、(F2)学术、(F3)表现(写作和音乐)、(F4)机械/科学、(F5) 艺术	0. 82-0. 87	EFA	I	A	VG	D
K-DOCS 中文版	涂翠平和樊富珏, 2015	自评	大学生	共 50 个题项; 采用李克特 5 点量表	5	(F1)日常、(F2)学术、(F3)表现、(F4)机械/科学、(F5) 艺术	0. 93	CFA	I	VG	VG	VG
SSCS (创造性自我简版量表)	Karwowski et al., 2018	自评	10 岁及以上学生	共 11 个题项; 采用李克特 5 点量表	3	(F1)创造性自我效能感(CSE)、(F2)创造性个人认同(CPI)、(F3)自我评定创造力	0. 83-0. 84	CFA+EFA	I	VG	VG	D
BICB(创造性行为自传调查)	Silvia et al., 2021	自评	大学生、成人	共 34 个题项; 采用二分法	1	单因子结构	0. 86	CFA+EFA; IRT/Rasch	I	VG	I	VG
RIBS (Runco 观念行为量表)	Runco et al., 2001	自评	大学生	共 23 个题项; 采用李克特 5 点量表	1/2	因双因子缺乏理论支撑, 解读结果时应以单因子结构为指导框架	0. 92	CFA+EFA	I	VG	VG	D
CAQ (创造力成就问卷)	Carson et al., 2005	自评	大学生	共 96 个题项; 分 3 部分进行勾选	3/2	三因子: (F1)表达型、(F2)科学型、(F3)表演型; 二因子: (F1)艺术因子、(F2)科学因子	0. 96	PCA(未做 EFA/CFA)	I	D	VG	D
CAQ 问卷中文修订版	王战旗和张兴利, 2020	自评	成人	共 96 个题项; 分 3 部分进行勾选	3/2	三因子: (F1)幽默、写作、舞台影视表演和视觉艺术、(F2)科学探究、建筑和发明、(F3)舞蹈和音乐; 二因子: (F1)艺术因子、(F2)科学因子(建筑、舞台)	NR	PCA(未做 EFA/CFA)	I	D	NR	D

内容效度是所有量表的共同短板。所有量表均未对目标人群进行系统的认知访谈以检验条目的可理解性,且多数缺乏完整的专家评审报告,因此全部评为“不充分”。结构效度方面,7个量表采用验证性因子分析或项目反应理论并报告完整拟合指标,评为“非常好”;K-DOCS 原版通过探索性因子分析获得清晰结构,评为“充分”;CAQ 原版和 CAQ 中文版仅用主成分分析,评为“不确定”。内部一致性方面,8个量表 Cronbach's α 均 ≥ 0.70 且具单维性,7个评为“非常好”,BICB 因条目局部依赖评为“不确定”,CAQ 中文版未报告。构念效度假设检验方面,K-DOCS 中文版和 BICB 同时提供收敛与区分效度充分证据,评为“非常好”;CMS、K-DOCS 原版、SSCS、RIBS、CAQ 原版、CAQ 中文版仅有收敛效度、缺区分效度,评为“不确定”;大学生创造力量表完全缺失收敛效度检验,评为“不充分”。

3.2. 测量属性质量评价 纳入量表的测量属性评价见表 2。在内容效度方面,所有量表均未对目标人群进行认知访谈或可理解性预测试,相关信息不足,全部评为“不确定”。这反映的是原始研究在内容效度验证上的报告缺失,而非量表条目本身存在质量问题。结构效度方面,K-DOCS 原版与 K-DOCS 中文版因满足标准,评为“充分”。CAQ 原版与 CAQ 中文版仅用主成分分析且累积方差未达标,评为“不充分”。其余量表因 EFA 累积方差不足或 CFA 拟合指标未达标,评为“不充分”。内部一致性方面,除 CAQ 中文版未报告,CMS 的增长型心态因子未达标而评为“不充分”外,其余量表均报告 Cronbach's $\alpha \geq 0.70$,且具备至少低证据的单维性,评为“充分”。构念效度假设检验方面,K-DOCS 中文版和 BICB 提出明确假设且结果一致性 $\geq 75\%$,评为“充分”。大学生创造力量表未进行任何假设检验,评为“不充分”。其余量表虽有部分收敛效度证据,但未提出明确假设或缺少区分效度检验,评为“不确定”。

表 2 纳入量表的测量属性质量评价

量表名称	结构效度关键数据	COSMIN 测量属性			
		CV	SV	IC	HV
大学生创造力量表	EFA 累积方差 =48.619%;CFA:CFI=0.907,RMSEA=0.063	?	-	+	-
CMS(创造力心态量表)	EFA 累积方差=50%; CFA:CFI=0.92,RMSEA=0.073	?	-	-	?
K-DOCS (Kaufman 领域创造力量表)	5 因子载荷 ≥ 0.45 ;碎石图支持 5 因子结构	?	+	+	?
K-DOCS 中文版	CFA: RMSEA=0.039	?	+	+	+
SSCS (创造性自我简版量表)	CFA:CFI=0.94,RMSEA=0.08	?	-	+	?
BICB(创造性行为自传调查)	存在 6 对局部依赖,违反局部独立性标准	?	-	+	+
RIBS (Runco 观念行为量表)	CFA:CFI=0.792,TLI=0.771,RMSEA=0.09	?	-	+	?
CAQ (创造力成就问卷)	PCA:三因子累积方差 43.8%; 双因子 33.5%	?	-	+	?
CAQ 问卷中文修订版	PCA:三因子累积方差 47.4%; 双因子 35.3%	?	-	NR	?

注: CV=内容效度, SV=结构效度, IC=内部一致性, HV=构念效度假设检验; +=充分, ?=不确定, -=不充分, NR=未报告

3.3. 证据质量分级 内容效度所有量表均为低级证据。结构效度中,K-DOCS 原版与 K-DOCS 中文版为高级证据,其余量表因明确方法学缺陷为低级。内部一致性中,除 CMS 为中级、CAQ 中文版为极低级外,其余均为高级证据。构念效度假设检验中,K-DOCS 中文版和 BICB 为高级证据,其余为低级或极低级。

4. 讨论

4.1. 量表选择建议 若研究目的为评估多领域创造力,K-DOCS 中文版是当前测量学属性最全面的选择,其在结构效度、内部一致性和构念效度方面均表现优异。若研究聚焦于创造性自我信念,SSCS 具有清晰的三因子结构和良好的内部一致性,但需注意其区分效度证据不足。若研究需要测量创造性行为或成就,BICB 是构念效度证据最充分的工具,尤其适合精细化或追踪研究;CAQ 中文版应用广泛,但因其计分方式特殊未报告内部一致性(以重测信度替代),

使用时需结合具体研究需求。大学生创造力量表在内容效度方面表现较好（有专家和学生参与条目生成），但构念效度检验缺失，适用于本土化探索性研究。

4.2. GenAI 辅助测评工具的发展与未来优化方向 GenAI 正推动创造力测评走向“人机协同”新范式，主要路径包括：基于大语言模型的自动评分（如 MOTES 系统、GPT-4o 评分），可实现对发散思维任务的客观高效评分；测评题目智能生成（如 CPIG 框架、大语言模型情境化命题），可大规模生成情境化命题；多模态智能测评（如图形发散思维自动评分），拓展了非言语创造力维度。未来亟需建立适用于 GenAI 辅助测评工具的质量评价框架，且应规范内容效度流程，增加对目标人群的认知访谈并报告内容效度指数，补全验证性因子分析验证并报告拟合指标，强化区分效度检验。

5. 结论

9 个创造力自评量表内部一致性良好，但内容效度普遍缺乏可理解性检验，构念效度验证不完整。K-DOCS 中文版和 BICB 的测量学属性最为全面。研究者应结合研究目的选用，并关注报告规范性。生成式人工智能辅助测评正成为重要补充，建议未来采用传统自评量表与 AI 辅助测评相结合的混合模式。

参考文献

- [1] 王战旗 & 张兴利.(2020).创造力成就问卷的中文修订.心理与行为研究,18(03),390-397.
- [2] 江怀.(2011).大学生创造力量表的初步编制与现状调查研究(硕士学位论文,福州大学).硕士
- [3] 何少颖,方沛,李洋,江怀 & 任志洪.(2015).大学生创造力量表编制.中国健康心理学杂志,23(01),122-125.
- [4] 涂翠平 & 樊富珉.(2015).Kaufman领域创造力量表中文版的验证及初步应用.心理与行为研究,13(06),811-816.
- [5] Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity research journal*, 17(1), 37-50.
- [6] Mokkink, L. B., Elsmans, E. B., & Terwee, C. B. (2024). COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Quality of Life Research*, 33(11), 2929-2939.
- [7] Runco, M. A., Plucker, J. A., & Lim, W. (2001). Development and psychometric integrity of a measure of ideational behavior. *Creativity Research Journal*, 13(3-4), 393-400.
- [8] Silvia, P. J., Rodriguez, R. M., Beaty, R. E., Frith, E., Kaufman, J. C., Lopera, P., & Reiter-Palmon, R. (2021). Measuring everyday creativity: A Rasch model analysis of the Biographical Inventory of Creative Behaviors (BICB) scale. *Thinking Skills and Creativity*, 39, 100797.
- [9] Kaufman, J. C. (2012). Counting the muses: development of the Kaufman domains of creativity scale (K-DOCS). *Psychology of Aesthetics, Creativity, and the Arts*, 6(4), 298.
- [10] Karwowski, M. (2014). Creative mindsets: Measurement, correlates, consequences. *Psychology of Aesthetics, Creativity, and the Arts*, 8(1), 62.
- [11] Karwowski, M., et al. (2018). Measuring creative self-efficacy and creative personal identity. *The International Journal of Creativity & Problem Solving*, 28(1), 45-57.