

游戏化认知测评的理论建构、全球经验与中国探索

严梓鸣, 赵春源

苏州科技大学教育学院, 江苏 苏州

收稿日期: 2026年5月17日; 录用日期: 2026年6月11日; 发布日期: 2026年6月29日

摘要

游戏化测评通过将游戏机制融入评估过程, 为破解传统测评参与度低、情境真实性不足等局限提供了新路径。本文系统梳理了该领域的理论演进、全球实证与中国本土实践。证据中心设计为游戏化测评的科学化开发提供了方法论框架, 自我决定理论解释了其用户中心优势。全球实证研究表明, 游戏化测评与传统认知测验呈中等程度相关, 在用户投入、焦虑缓解与评估公平性方面具有优势, 但其效度受游戏设计品质多重调节。中国本土研究已从理论译介走向初步工具研发, 但仍面临常模缺失与跨文化效度证据不足的瓶颈。未来需着力构建标准化常模体系, 发展AI驱动自适应测评系统, 并深化非认知能力的游戏化测量探索。

关键词

游戏化测评, 认知能力, 证据中心设计, 常模建设, 机器学习

Game-Based Cognitive Assessment: Theoretical Construction, Global Evidence and Chinese Exploration

Ziming Yan, Chunyuan Zhao

School of Education, Suzhou University of Science and Technology, Suzhou Jiangsu

Received: May 17, 2026; accepted: June 11, 2026; published: June 29, 2026

Abstract

Game-based assessment, by integrating game mechanics into evaluation processes, offers a novel approach to overcoming the limitations of traditional assessment, such as low participant engagement

and insufficient ecological validity. This paper systematically reviews the theoretical evolution, global empirical evidence, and Chinese indigenous practices in this field. Evidence-centered design provides the methodological framework for the scientific development of game-based assessment, while self-determination theory explains its user-centered advantages. Global empirical studies indicate a moderate correlation between game-based and traditional cognitive tests, with game-based approaches demonstrating advantages in user engagement, test anxiety reduction, and assessment fairness, though their validity is moderated by game design quality. Chinese indigenous research has progressed from theoretical introduction to initial tool development, yet still faces bottlenecks including the absence of standardized norms and insufficient cross-cultural validity evidence. Future efforts should focus on constructing standardized norm systems, developing AI-driven adaptive assessment systems, and deepening the exploration of game-based measurement of non-cognitive abilities.

Keywords

Game-Based Assessment, Cognitive Ability, Evidence-Centered Design, Norm Construction, Machine Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

认知能力作为个体学习、工作和适应环境的核心心理资源,其科学评估在教育诊断、职业选拔及临床干预中具有基础性地位。过去一个世纪,心理测量学围绕智力结构理论发展出了系统的标准化纸笔测验工具,这些工具在信效度指标和常模参照体系方面达到了高度成熟。然而,随着数字社会的深度转型和评估场景的日趋多元,传统测评范式的固有局限日益凸显:静态的题目呈现方式缺乏情境真实性,难以捕捉个体在动态复杂环境中的认知加工过程;重复性答题模式导致受测者参与动机不足,测验焦虑、社会赞许性偏差可能污染测量结果;传统计分方式仅记录最终作答,大量反映认知策略和加工效率的过程信息被遗失。

游戏化测评(Game-Based Assessment)的兴起为突破上述困境提供了可能。游戏化测评是指通过游戏或游戏化活动,对个体的能力、人格等心理特性和行为进行量化测评的方法(徐俊怡,李中权,2020)。与传统测评相比,游戏化测评利用沉浸式交互环境、多模态数据采集和自适应反馈调控等技术优势,能够在更接近真实情境的条件下激活受测者的认知潜能,并通过细粒度的行为日志记录实现从“结果评价”到“过程刻画”的跃升。已有研究证实,游戏化任务可有效提升认知评估的信效度(Anguera et al., 2013)。

近年来,游戏化测评在学术研究和商业实践中均呈高速发展态势。企业在招聘领域已率先尝试将游戏测评工具投入应用,如普华永道开发的 Career Unlocked 等招聘测评游戏已投入商业运营。然而,与蓬勃发展的实践相比,游戏化测评的理论建构和学术验证仍相对滞后(徐俊怡,李中权,2020)。学界对游戏化测评的有效性仍存在争议:部分研究验证了其与传统测验的高度相关(Landers et al., 2022),另一些研究则指出游戏机制可能引入与目标构念无关的认知负荷,损害测量的纯净性(Wu et al., 2022)。

本文旨在通过对国内外研究文献的系统梳理与批判性分析,厘清游戏化认知测评的理论基础、技术框架与全球实证证据,评估其在教育、临床及企业等场域的实践进展,并聚焦中国本土研究的发展现状与独特挑战。

2. 理论根基：从游戏机制到测评设计的逻辑闭环

2.1. 证据中心设计：游戏化测评的方法论基石

游戏化测评要实现从“有趣的的游戏”到“有效的测评”的跨越，不能仅凭经验将游戏与心理测量简单叠加，而需构建严密的理论与技术闭环。证据中心设计框架为此提供了核心方法论。该框架包括三个核心组件：能力模型定义测评的目标构念及其结构维度；任务模型设计引发目标行为的游戏情境与交互方式；证据模型建立从可观察的游戏行为到不可观察的心理特质的推断逻辑，包括统计规则和统计模型(徐俊怡, 李中权, 2020)。这三个组件构成“目标构念→设计任务→收集证据→推断能力”的完整推理链，是游戏化测评区别于随意设计游戏的关键所在。

Landers 和 Sanchez (2022)将游戏相关评估划分为三种范式：基于游戏的评估(一种独立的评估方法)、游戏化设计评估(从零开始融入游戏概念的设计策略)和游戏化评估(向现有评估添加游戏机制的重新设计策略)。这一分类在 Ramos-Villagrasa 等人(2022)的系统综述中得到进一步发展，形成了一个从“传统评估”到“娱乐性游戏”的连续谱系。理解这种类型学差异对于评估不同游戏化测评工具的适用范围具有重要意义。

2.2. 心理需求满足与心流体验：游戏化的动机机制

游戏化测评之所以能比传统测验更有效地激活个体认知潜能，深层原因在于其满足人类基本心理需求的独特能力。Suh 等人(2018)基于认知评价理论构建了用户投入度模型，并在游戏化知识共享平台的实际用户数据上得到验证：游戏动力学特征(奖励、竞争、利他、自我表达)通过满足自主性、胜任力和关系性需求，间接提升用户的愉悦感和投入度。

心流理论进一步解释了游戏化任务中注意力高度集中的心理机制。当游戏难度过低时玩家感到无聊，难度过高则引发焦虑。游戏化测评可通过算法动态调整任务参数，使受测者持续处于“挑战-技能”平衡区间，既保证测评精确性，又维持用户体验流畅性。Molina-Torres 等人(2021)对物理治疗专业学生的研究发现，游戏化教学组不仅期末成绩显著优于传统讲授组，学习满意度也提升了 37%，这正是心理需求满足促进学习投入的实证体现。

2.3. 从浅层激励到深层设计：理论的澄清与演进

早期游戏化实践多采用积分、徽章、排行榜等浅层激励机制。随着实践深入，研究者逐渐认识到，仅仅叠加外在奖励难以产生持久的行为改变，甚至可能削弱个体对任务的内在兴趣。为此，学者提出了“游戏化设计”概念，强调从追求外在激励框架转向创造内在满足的心理体验(Dichev et al., 2015)。这一转变将设计关注点从“如何让用户赢取积分”转移到“如何让用户在完成任务时获得真实的成长体验和胜任感”。在游戏化测评领域，这尤为重要——相比于依赖外在奖励驱动用户完成更多测评题目，通过巧妙的任务设计和叙事建构激发用户内在动机，更有利于获得反映用户真实能力的测评结果。

3. 全球经验：游戏化测评的实证证据与边界条件

全球实证研究在过去十年中呈现快速增长趋势，研究者从认知能力评估、非认知能力评估、教育学习和职业培训等多元场域出发，积累了丰富证据。

3.1. 认知能力评估：效度证据与调节因素

认知能力是游戏化测评研究最活跃的领域之一。Bipp 等人(2024)的元分析整合了 44 篇文献、52 个独立样本(总计超过 6100 名成人参与者)数据，发现游戏相关评估与传统认知能力测验的总体观测相关系数

为 $r=0.30$ (经信度校正后 $r=0.45$), 表明游戏表现确实反映了认知能力, 但二者并非简单等同。测量范围的广度显著调节两者关系: 当传统认知能力以多个分测验聚合分数表达, 或游戏化测评分数来自多个游戏任务时, 相关性显著强于单一指标情况。但出乎意料的是, 游戏化测评的测量目的(是否专门设计用于认知测评)并未显著调节效度——即使是娱乐目的开发的商业休闲游戏, 其表现也与认知能力存在显著关联(Bipp et al., 2024)。

在具体研究案例方面, Landers 等人(2022)基于理论驱动设计框架开发的 Cognify 游戏化认知能力测验套件, 展现出与传统测验高度聚合的优良特性, 潜变量相关高达 0.97。Peters 等人(2021)在三维沉浸式 Minecraft 环境中实现了图形完成、心理旋转和空间建构三项认知思维任务, 研究发现这些任务与传统纸笔测验具有良好的聚合效度(潜变量相关 0.72), 更开创性地揭示游戏日志中的行为特征(如垂直位置移动范围、移动路径效率)可显著预测空间推理能力, 充分展现了过程数据在认知评估中的独特价值。

然而, 并非所有游戏化测评都有理想结果。Ohlms 等人(2024)基于 Minecraft 开发的冒险类游戏化测评, 虽与传统纸笔测验存在较强相关($r=0.51$), 但在申请者反应的所有维度上均显著逊色于传统测验——这一发现打破了“游戏化等于更好用户体验”的常见假设。此外, Wu 等人(2022)的研究提供了另一个警示案例: 一款旨在测量尽责性维度的游戏化任务, 其表现实际与认知能力的相关性高于与尽责性本身的相关性。这种“测 A 结果量到 B”的内容污染问题, 提示研究者在开发阶段就必须进行严格的构念效度验证。

3.2. 非认知能力评估: 人格特质与社会情绪

相较于认知能力, 非认知能力因更易受社会赞许性影响, 其有效测量一直是心理测评领域的难题。游戏化测评因具有情境隐蔽性——受测者不易觉察被评估的具体构念——在这方面展现出独特优势。

Ramos-Villagrasa 等人(2024)对传统大五人格量表进行了叙事性游戏化改编(VASSIP), 在保留原有题目内核的同时融入科幻故事线、环境音乐和沉浸式图像。结果表明, 游戏化版本在信度和对任务绩效、反生产行为的预测上与传统量表相似, 但在组织吸引力、舒适度和预测效度感知方面获得了显著更积极的用户评价。然而, 游戏化版本的尽责性分数系统偏高, 提示游戏化改编可能引入了与目标构念无关的变异。

此外, 基于游戏的测评在评估社会情绪能力方面也取得积极结果。DeRosier 和 Thomas (2018)开发的 ZooU 社会情绪能力评估游戏通过虚拟社交场景测量儿童的交流、合作、同理心等能力, 游戏得分与传统教师评定量表呈显著相关, 初步支持了游戏化测评在非认知能力筛查中的临床潜力。

3.3. 学习与培训领域: 反馈机制的关键作用

教育领域是游戏化测评研究最丰富的应用场景之一。Zainuddin 等人(2020)比较了三种游戏化电子测验工具与传统纸质测验在科学课上的作用, 发现游戏化测验更能激发学生情感、行为、认知和能动性四个维度的学习投入, 但两组在知识掌握程度上并无显著差异。这一结果暗示游戏化测评在提升学习过程质量方面的优势, 可能比其直接影响知识习得更为突出。

反馈类型是影响游戏化测评教育效果的关键设计要素。Tsai 等人(2015)在井字棋游戏化测评平台上发现, 提供即时精细反馈的实验组, 无论游戏模式是单人还是多人, 其知识获取效率和游戏技能提升都显著优于无即时反馈组。Yang 和 Lu (2021)的研究进一步发现, 基于二阶测试框架诊断学生数学错误概念的游戏, 当学生愿意花时间阅读错误概念反馈时, 其学习效果和数学焦虑缓解都明显优于对照组。这些研究指向一个核心原则: 游戏化测评的效果高度依赖于其内嵌的教学诊断与反馈机制的质量。

在职业培训领域, Tuti 等人(2020)在肯尼亚等资源匮乏地区开展的智能手机游戏化医疗急救培训实验

发现, 自适应反馈机制在控制学习者背景差异后仍能带来显著的学习增益, 且即时重复训练模式优于间隔一周后的训练模式, 为在资源不均衡区域推广低成本培训提供了创新思路。

3.4. 用户投入度与评估公平性

用户投入度是游戏化测评区别于传统测验的核心优势之一。Suh 等人(2018)的研究验证了游戏动力学特征通过满足基本心理需求间接提升用户愉悦感和投入度的机制模型。

评估公平性是游戏化测评在社会功能维度的重要议题。传统认知能力测验因存在显著的子群体差异而在招聘应用中面临法律和伦理挑战。Leutner 等人(2023)在超过一万名真实求职者中验证了机器学习优化后的游戏化认知能力评估系统, 创新性地模型训练过程中引入偏差惩罚项, 在保证预测效度的同时有效缩小子群体差异。该系统的净推荐值达到 58, 表明大多数求职者对其持积极态度。不过, 负面反馈主要集中在“工作相关性”疑虑上, 提示游戏化测评实践需在设计阶段就将沟通策略纳入考量。

4. 方法前沿: 从结果数据到过程数据的分析跃迁

4.1. 过程数据的挖掘与行为特征工程

游戏化测评相较于传统测验的一大显著优势, 在于其能够记录受测者与游戏系统交互的全部过程数据。Peters 等人(2021)在 Minecraft 测评研究中, 通过特征重要性排序和局部加权回归分析, 识别出高效能玩家在空间建构任务中的典型行为特征: 他们在垂直维度上有更大活动范围, 移动路径更高效, 与目标对象保持适当的观察距离。这些特征是空间推理策略差异在游戏行为中的系统性映射。

然而, 目前对过程数据的利用率仍有限。de Klerk 等人(2015)总结了 31 项游戏化测评相关研究成果, 发现其中只有 6 项研究单独使用了过程数据, 10 项研究仅使用结果数据, 其余研究虽两者兼用, 但往往将过程数据简单降维为少量汇总指标。如何从海量日志数据中提取具有心理测量意义的有效特征, 仍是方法学研究的前沿课题。Alonso-Fernández 等人(2019)提出的游戏学习分析框架, 通过标准化 xAPI-SG 数据格式收集游戏内各类交互操作, 为过程数据的系统化采集提供了可推广的技术方案。

4.2. 机器学习驱动的评分模型与自适应测评

机器学习的引入为游戏化测评的评分环节提供了更强大的建模手段。Leutner 等人(2023)的研究证明, 在模型训练过程中引入偏差惩罚项, 可在维持预测准确性($r=0.50$)的同时, 显著缩小不同性别、年龄、种族子群体之间的评估差异, 为游戏化测评在招聘选拔等高风险场景下实现“高精准度 - 高公平性”双重目标提供了技术路径。

自适应测评系统是另一值得关注的探索方向。这类系统通过持续分析受测者行为数据实时估计能力水平, 并据此动态调整后关卡难度和内容, 使每位受测者面对的任务挑战始终与当前能力匹配, 在整个测评过程中维持心流状态。

4.3. 信效度检验的特殊考量

游戏化测评的信效度检验继承传统心理测量学基本框架, 但面临独特挑战。在信度评估上, 传统内部一致性系数在游戏化情境中可能不完全适用——游戏关卡往往难以被视为平行题目的简单重复样本, 且关卡间存在学习迁移效应。研究者提出以重测信度作为替代方案(Peters et al., 2021), 但因存在练习效应, 如何校准重复测量带来的分数膨胀亦需解决。

效度验证方面, 目前文献主要围绕构念效度、效标关联效度和增量效度三条路径展开。Ramos-Villagrasa 等人(2022)的系统综述指出, 现有游戏相关评估的效应量大多集中在 $r=0.20$ 至 0.50 的中低至

中高范围, 表明游戏化测评与传统测验测量的是相关但不同的构念。All 等人(2021)在数字游戏学习有效性评估研究中特别强调研究设计严谨性对内部效度的影响, 指出随机化失败和前测效应是主要威胁因素, 建议采用区组随机化和经过预试的平行版本测验加以控制。

5. 中国探索：本土实践的发展现状与独特挑战

5.1. 理论译介与自主创新

游戏化测评在国内的研究起步较晚, 但发展势头迅猛。徐俊怡和李中权(2020)是国内较早系统介绍游戏化测评概念与范式的学者, 详细阐述了其定义、证据中心设计测评范式及在认知和非认知能力评估中的应用前景。

在自主理论创新方面, 黄凯奇等人(2025)提出了“基于交互环境的智能心理测评”概念, 从测评需求、测评环境和分析技术三个核心要素重新定义了游戏化测评的系统架构。该研究以 Minecraft 游戏和自主研发的箱庭疗法电子化系统《AI 心世界》为案例进行实证展示, 其中《AI 心世界》在强迫、抑郁和焦虑维度上的评估结果与 SCL-90 量表得分存在显著相关(相关系数分别为 0.589、0.598、0.665), 初步支持了 AI 驱动的游戏化测评工具在非认知能力评估中的临床有效性。

5.2. 教育领域的初步实践

在游戏化学习与测评的结合方面, 国内研究积累了一些初步实证发现。Zhao 等人(2024)将主题游戏式学习融入精神暴力风险评估与应对培训, 随机对照实验结果表明, 游戏组护生在暴力风险评估与应对方面的技能水平和自信心明显优于情境模拟组。Yang 和 Lu (2021)对中国学生数学错误概念的游戏化诊断研究则发现, 重视阅读错误反馈的学生的数学学习效果和焦虑缓解均显著优于对照组。这些研究提示, 游戏化测评的本土化需要深入研究本土学习者的认知模式、动机特征和评估偏好。

5.3. 挑战与进路

尽管中国游戏化测评研究已取得一定进展, 但目前仍面临几个关键挑战。其一, 标准化常模体系的缺失。已被验证有效的游戏化测评工具大多是在西方样本中完成常模研究的, 缺乏基于中国本土大规模人群的常模参照, 使得测评结果无法在统一框架中进行标准化比较。其二, 自主研发的本土化测评工具有限, 大量研究仍处于引进消化阶段。其三, 跨文化效度比较研究几乎空白——西方验证有效的游戏任务直接移植到中国样本时是否会因文化差异产生系统偏差, 目前仍知之甚少。

6. 结论与展望

本文对游戏化认知测评领域的理论演进、全球实证证据与中国本土实践进行了系统梳理。从理论基础看, 证据中心设计为游戏化测评的科学化开发提供了严密的方法论框架; 从实证证据看, 游戏化测评在构念效度上与传统纸笔测验呈中等至中高度相关; 从中国实践看, 本土研究已从理论译介走向初步的自主工具研发。在全面评估这一新兴领域时, 有必要对其独特优势与潜在风险进行平衡审视。

6.1. 独特优势

综合现有实证证据, 游戏化测评相较传统方法展现出四个维度的差异化优势。

在用户体验维度, 游戏化测评通过满足自主性、胜任力和关系性等基本心理需求(Suh et al., 2018), 显著提升了受测者的参与动机与测试体验。Molina-Torres 等人(2021)的对照实验表明, 游戏化教学组的学习满意度较传统组提升了 37%。这种体验优势在数字原住民群体中尤为突出, 有望缓解传统测评中普遍存在的测验焦虑和参与度不足问题。

在数据维度, 游戏化测评实现了从“结果导向”到“过程刻画”的范式跃迁。Peters 等人(2021)的研究证明, 通过分析游戏日志中的空间移动轨迹、操作序列等行为特征, 可有效预测受测者的空间推理能力, 而这些过程信息是传统纸笔测验无法获取的。Alonso-Fernández 等人(2019)提出的标准化数据采集方案, 为过程数据的系统化利用提供了可推广的技术框架。

在评估维度, 游戏化测评的情境隐蔽性有助于降低社会赞许性偏差对非认知能力评估的影响。Ramos-Villagrasa 等人(2024)的研究发现, 游戏化人格测评在组织吸引力等指标上获得了比传统量表更积极的用户反馈。在认知能力评估领域, Landers 等人(2022)的理论驱动设计框架实现了游戏化任务与传统测验的高水平聚合(潜变量相关 0.97), 为构念效度提供了有力证据。

在公平性维度, Leutner 等人(2023)的研究初步证明, 通过在算法层面引入偏差惩罚机制, 游戏化测评可在保证预测效度的同时有效缩小子群体差异, 为破解传统认知测验长期面临的不利影响问题提供了技术路径。

6.2. 潜在风险与伦理挑战

在充分肯定游戏化测评优势的同时, 研究者也需正视其潜在的负面影响与伦理风险。

其一, 数字鸿沟与公平性悖论。虽然游戏化测评有望通过算法优化缩小子群体差异(Leutner et al., 2023), 但游戏经验的个体差异可能引入新的偏差。已有研究表明, 男性及有视频游戏经验的受测者在游戏化测评中的表现通常略优(Ohlms et al., 2024; Bipp et al., 2024), 这提示游戏化形式本身可能对不同数字素养水平的群体产生差异化影响。若缺乏对游戏经验的系统控制, 游戏化测评可能在弥合传统评估偏差的同时制造新的不公平。

其二, 数据隐私与算法透明度。游戏化测评通过过程数据采集获取的受测者行为信息(如反应时间、操作序列、空间移动轨迹)远超传统测评, 这些细粒度数据在提升评估精度的同时, 也带来了隐私泄露风险。如何界定“必要数据”与“过度采集”的边界, 如何向受测者清晰说明数据用途并获取有效知情同意, 是亟需解决的伦理问题。此外, 当机器学习算法参与评分决策时, 算法的可解释性不足可能导致“黑箱评估”, 使受测者难以理解或质疑评估结果。

其三, 认知负荷与过度沉浸的平衡。游戏化设计旨在提升用户投入度, 但过度沉浸可能导致受测者忽视任务本身的认知要求, 或因长时间游戏产生认知疲劳。不当的游戏化设计可能增加无关认知负荷而损害任务表现。如何在激发参与动机与控制认知负荷之间取得平衡, 需要更精细的设计准则和实证指导。

其四, 开发与维护的成本效益考量。相较于传统纸笔测验, 高质量游戏化测评的开发需要跨学科团队协作, 投入更高昂的时间与资金成本(All et al., 2021), 且后期维护(内容更新、平台适配、安全防护)同样需要持续资源投入。在资源有限的场景(如基层学校、社区筛查)中, 高昂成本可能限制其推广应用。

6.3. 未来具体研究方向

基于上述分析, 未来研究可围绕以下具体问题展开。

在工具开发与验证方面, 首先需优先推进中国本土常模建设。可选取 2~3 个已获初步效度证据的游戏化认知任务(如 Minecraft 环境下的空间推理任务), 在涵盖不同年龄段、城乡背景和教育水平的大样本中进行标准化施测, 建立分群体参照常模。其次, 应系统开展跨游戏比较研究: 针对同一认知构念(如工作记忆), 设计多个不同游戏机制的测评版本, 通过多质多法矩阵比较其聚合效度与区分效度, 以分离“构念本身”与“游戏形式”的效应。再次, 需建立游戏化测评的效度验证标准框架, 明确不同应用场景(筛查、诊断、选拔)所需的最低效度阈值和信度指标要求。

在技术优化方面, 应重点探索轻量级自适应算法的应用。现有自适应系统多依赖大规模训练数据,

而实际场景中数据往往有限。研究可比较不同自适应策略(如贝叶斯知识追踪、强化学习)在小样本条件下的表现,寻找效率与精度之间的最优平衡。此外,需研发可解释的机器学习评分模型,使评估结果不仅能给出分数,还能回溯性地解释哪些行为特征支持了该分数判断,以增强算法透明度和用户信任。

在伦理与公平性保障方面,应建立游戏化测评的公平性审计框架,系统检验游戏经验、年龄、性别、文化背景等因素对不同测评任务的差异化影响,并开发相应的偏差校正方法。同时,需研究数据最小化采集策略,在保证评估精度的前提下,尽可能减少对受测者隐私信息的获取。

在非认知能力评估方面,当前研究仍处于初级阶段。未来可重点探索利用游戏中的社会互动情境测量共情、合作倾向、道德判断等传统量表难以有效评估的构念。例如,设计包含虚拟角色交互的叙事游戏,通过分析受测者的对话选择和资源分配行为推断其社会情绪能力。这类研究的核心挑战在于建立从复杂社交行为到稳定人格特质的有效证据链。

游戏化认知测评的长远发展,需要在技术创新与伦理审慎之间寻求平衡。研究者既不能因潜在风险而回避技术革新,也不能因追求效率而忽视公平与隐私底线。只有在科学严谨性与社会责任感并重的前提下,游戏化测评才能真正成为服务教育、临床和职业领域的可靠工具。

基金项目

此论文为江苏省心理学会 2025 年度本科生研究基金暨创新实践项目一般项目的内容之一,受江苏省心理学会资助,项目编号 JSPSY2025YB020。

参考文献

- 黄凯奇, 康雅萱, 晏成信, 胡世宇, 高文斌, 王利刚, 陶婷, 曾杨睿宇(2025). 基于交互环境的智能化心理测评(综述). *中国心理卫生杂志*, 39(4), 337-343.
- 徐俊怡, 李中权(2020). 基于游戏的心理测评. *心理科学进展*, 29(3), 394-403.
- All, A., Castellar, E. N. P., & Van Looy, J. (2021). Digital Game-Based Learning Effectiveness Assessment: Reflections on Study Design. *Computers & Education*, 167, Article ID: 104160. <https://doi.org/10.1016/j.compedu.2021.104160>
- Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Lessons Learned Applying Learning Analytics to Assess Serious Games. *Computers in Human Behavior*, 99, 301-309. <https://doi.org/10.1016/j.chb.2019.05.036>
- Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J. et al. (2013). Video Game Training Enhances Cognitive Control in Older Adults. *Nature*, 501, 97-101. <https://doi.org/10.1038/nature12486>
- Bipp, T., Wee, S., Walczok, M., & Hansal, L. (2024). The Relationship between Game-Related Assessment and Traditional Measures of Cognitive Ability—A Meta-Analysis. *Journal of Intelligence*, 12, Article 129. <https://doi.org/10.3390/jintelligence12120129>
- de Klerk, S., Veldkamp, B. P., & Eggen, T. J. H. M. (2015). Psychometric Analysis of the Performance Data of Simulation-Based Assessment: A Systematic Review and a Bayesian Network Example. *Computers & Education*, 85, 23-34. <https://doi.org/10.1016/j.compedu.2014.12.020>
- DeRosier, M. E., & Thomas, J. M. (2019). Hall of Heroes: A Digital Game for Social Skills Training with Young Adolescents. *International Journal of Computer Games Technology*, 2019, Article 6981698. <https://doi.org/10.1155/2019/6981698>
- Dichev, C., Dicheva, D., Angelova, G., & Agre, G. (2015). From Gamification to Gameful Design and Gameful Experience in Learning. *Cybernetics and Information Technologies*, 14, 80-100. <https://doi.org/10.1515/cait-2014-0007>
- Landers, R. N., & Sanchez, D. R. (2022). Game-Based, Gamified, and Gamefully Designed Assessments for Employee Selection: Definitions, Distinctions, Design, and Validation. *International Journal of Selection and Assessment*, 30, 1-13. <https://doi.org/10.1111/ijasa.12376>
- Landers, R. N., Armstrong, M. B., Collmus, A. B., Mujcic, S., & Blaik, J. (2022). Theory-Driven Game-Based Assessment of General Cognitive Ability: Design Theory, Measurement, Prediction of Performance, and Test Fairness. *Journal of Applied Psychology*, 107, 1655-1677. <https://doi.org/10.1037/apl0000954>
- Leutner, F., Codreanu, S., Brink, S., & Bitsakis, T. (2023). Game Based Assessments of Cognitive Ability in Recruitment:

- Validity, Fairness and Test-Taking Experience. *Frontiers in Psychology*, 13, Article 942662. <https://doi.org/10.3389/fpsyg.2022.942662>
- Molina-Torres, G., Rodríguez-Arrastia, M., Alarcón, R., Sánchez-Labraca, N., Sánchez-Joya, M., Roman, P. et al. (2021). Game-Based Learning Outcomes among Physiotherapy Students: Comparative Study. *JMIR Serious Games*, 9, e26007. <https://doi.org/10.2196/26007>
- Ohlms, M. L., Melchers, K. G., & Kanning, U. P. (2024). Can We Playfully Measure Cognitive Ability? Construct-Related Validity and Applicant Reactions. *International Journal of Selection and Assessment*, 32, 91-107. <https://doi.org/10.1111/ijasa.12450>
- Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and Validation of a Game-Based Intelligence Assessment in Minecraft. *Computers in Human Behavior*, 119, Article ID: 106701. <https://doi.org/10.1016/j.chb.2021.106701>
- Ramos-Villagrasa, P. J., Fernández-del-Río, E., & Castro, Á. (2022). Game-Related Assessments for Personnel Selection: A Systematic Review. *Frontiers in Psychology*, 13, Article 952002. <https://doi.org/10.3389/fpsyg.2022.952002>
- Ramos-Villagrasa, P. J., Fernández-del-Río, E., Hermoso, R., & Cebrián, J. (2024). Are Serious Games an Alternative to Traditional Personality Questionnaires? Initial Analysis of a Gamified Assessment. *PLOS ONE*, 19, e0302429. <https://doi.org/10.1371/journal.pone.0302429>
- Suh, A., Wagner, C., & Liu, L. (2018). Enhancing User Engagement through Gamification. *Journal of Computer Information Systems*, 58, 204-213. <https://doi.org/10.1080/08874417.2016.1229143>
- Tsai, F., Tsai, C., & Lin, K. (2015). The Evaluation of Different Gaming Modes and Feedback Types on Game-Based Formative Assessment in an Online Learning Environment. *Computers & Education*, 81, 259-269. <https://doi.org/10.1016/j.compedu.2014.10.013>
- Tuti, T., Winters, N., Edgcombe, H., Muinga, N., Wanyama, C., English, M. et al. (2020). Evaluation of Adaptive Feedback in a Smartphone-Based Game on Health Care Providers' Learning Gain: Randomized Controlled Trial. *Journal of Medical Internet Research*, 22, e17100. <https://doi.org/10.2196/17100>
- Wu, F. Y., Mulfinger, E., Alexander, L., Sinclair, A. L., McCloy, R. A., & Oswald, F. L. (2022). Individual Differences at Play: An Investigation into Measuring Big Five Personality Facets with Game-Based Assessments. *International Journal of Selection and Assessment*, 30, 62-81. <https://doi.org/10.1111/ijasa.12360>
- Yang, K., & Lu, B. (2021). Towards the Successful Game-Based Learning: Detection and Feedback to Misconceptions Is the Key. *Computers & Education*, 160, Article ID: 104033. <https://doi.org/10.1016/j.compedu.2020.104033>
- Zainuddin, Z., Shujahat, M., Haruna, H., & Chu, S. K. W. (2020). The Role of Gamified E-Quizzes on Student Learning and Engagement: An Interactive Gamification Solution for a Formative Assessment System. *Computers & Education*, 145, Article ID: 103729. <https://doi.org/10.1016/j.compedu.2019.103729>
- Zhao, J., Zeng, G., Chen, X., Huang, J., Xia, Z., Liang, R. et al. (2024). Effect of Theme Game-Based Learning on Psychiatric Violence Risk Assessment and Response Knowledge, Skills, and Self-Confidence of Nursing Students: A Randomized Controlled Trial. *Nurse Education Today*, 144, Article ID: 106419. <https://doi.org/10.1016/j.nedt.2024.106419>