

• 研究构想(Conceptual Framework) •

基于视频游戏的空间能力测评*

尚俊杰 石 祝 沈科杰

(北京大学教育学院学习科学实验室, 北京 100871)

摘 要 空间能力是个体对客体或空间图形在头脑中进行识别、编码、贮存、表征、分解组合和抽象概括的能力,是个体理解自身所处环境并解决问题的认知基础。准确、便捷、有效地测评空间能力,对增强 STEM 教育教学水平和人才培养质量都具有重要意义。由于空间能力受多因素共同作用,具有复杂性、多维度、内隐性的特点,使得利用计算机评价空间能力比较困难。本研究旨在准确、有效、大规模地测评空间能力,将使用多模态学习分析方法探索学习者空间认知行为表现特征,并基于视频游戏环境研发空间能力隐形测评关键技术与工具。具体包括:1)构建空间能力内在表征框架和评价指标体系;2)基于多模态学习分析构建学习者空间能力行为表现模型;3)探索视频游戏影响空间能力的关键因素,并使用游戏引擎开发基于视频游戏的测评工具;4)使用以证据为中心的设计框架和贝叶斯网络模型,开发并部署能够推断和预测空间能力的测评算法;5)在实验室和真实课堂情境开展实证研究,验证测评工具有效性。研究成果将有利于理解人类空间认知过程和行为表现,拓展和丰富空间能力相关理论,并为大规模数字化测评提供关键技术支持。

关键词 基于游戏的测评,空间能力,多模态学习分析,游戏化学习,隐形测评

分类号 B841

1 问题提出

空间能力(spatial ability),也称空间认知能力,是探测、处理、整合和构想环境中所有空间因素的能力,主要包括空间观察、空间记忆、空间思维和空间想象等方面(李洪玉,林崇德,2005)。空间能力是人和动物赖以生存的基础,个体在脑中对所有感官模态的空间信息和内在心理图像进行加工,从而理解自身所处的环境并解决各种问题(石祝,尚俊杰,2024)。空间能力水平将直接影响人们对科学、技术、工程、数学(science, technology, engineering, mathematics, 简称 STEM)领域问题的理解、表征和解决(Sorby & Panther, 2020; Uttal & Cohen, 2012),对于学习者的学业表现和职业选择具有重要意义。研究显示,空间能力与算术、

阅读和科学论述能力紧密相关,并能显著预测学习者的学习成绩(Geer et al., 2019; Hodgkiss et al., 2018; Liu et al., 2021; Xie et al. 2020)。大规模纵向追踪研究表明,空间能力对创造力发展也具有独特作用(Kell et al., 2013),而且能够显著预测青少年未来在 STEM 领域中所取得的职业成就(Wai et al., 2009)。因此,有针对性地培养和评价空间能力,不仅有助于提高学生的问题解决能力,也有利于在科技创新、航空航海等领域选拔人才(晏碧华, 2021; 周加仙, 蔡永华, 2013)。但由于空间能力的复杂性、多维度 and 内隐性,一直以来,对于空间能力的测量和评价都是困扰研究者和教育者的难题。这一方面导致了教育系统在选拔人才时忽视了空间能力的重要价值(Lubinski, 2010),另一方面也限制了空间能力理论研究和教学实践的发展。因此,探明空间能力的关键构成要素,构建空间能力评价指标体系,并基于真实情境开发准确、有效、易用的空间能力评价工具成为国内外研究者广泛关注的重要问题。

空间能力测评研究通常采用经典纸笔量表

收稿日期: 2024-03-28

* 国家自然科学基金“基于视频游戏的空间能力测评关键技术及工具开发研究”(62377001)资助。

通信作者: 石祝, E-mail: stonecolumnn@163.com

(如心理旋转测试),但由于研究者对空间能力内涵及其组成部分的不同定义,导致空间能力纸笔测试种类繁多、维度各异(Martín-Dorta et al., 2008)。而且,部分测试脱离真实情境,影响测评工具的生态效度,降低了对实际空间问题解决水平的预测性(Açikgöl et al., 2023)。在应用过程中,研究者也发现传统纸笔测试参与度较低,低年龄被试者难以完成较长的测试任务,并且测试过程中的霍桑效应会导致测试结果与预期偏离(石祝, 2019)。在此基础上,研究者也开始探索使用计算机化测评工具评价空间能力。早期的计算机化空间能力测评通常是纸笔测试迁移到计算机平台,其理论基础和评价量规都与传统纸笔测试大致相同。例如,心理旋转测试、空间定向测试、普渡空间可视化测试都有其计算机化版本,并且信度和效度都经过验证(Friedman et al., 2020)。虽然这类计算机化测试相比传统纸笔测试,具有简化评分程序,拓展测试人数的优点,还可以设置任务内容和任务流程,对数据加以收集和诊断(Hunt & Pellegrino, 1985),但由于受限于纸笔测试的研究框架,存在脱离真实情境、变量维度单一、缺乏过程性数据等缺点,并不能充分利用计算机在内容呈现、数据处理和自动化分析上的优势。

近年来,基于游戏的测评逐渐成为教育测量领域的热点(Hong & Liu, 2022; 徐俊怡, 李中权, 2021)。相比传统的计算机化测评,基于视频游戏的测评具有诸多优势。一方面,视频游戏具有多感官输入、沉浸式环境、智能化反馈等特点,所构建的虚拟环境能够有效促进复杂问题解决,并记录过程性数据来反映学习者行为(Owen & Baker, 2020)。学习者在游戏中的得分、运动轨迹、注视位置和时间等,都能作为评估空间能力的实证数据。另一方面,视频游戏能够创建贴近真实任务的虚拟场景,在游戏过程中嵌入无干扰的隐形测评(stealth assessment)能有效避免霍桑效应,增强测评的生态效度(Kim & Shute, 2015)。最重要的是,视频游戏是深受大众喜爱的互动媒体,它们可通过在线平台广泛发布,吸引了众多参与者,从而生成大规模的数据样本(Spiers et al., 2023),为面向广大人群的认知能力测试和大数据研究奠定基础(Xu et al., 2023)。由此可见,视频游戏为认知能力的计算机化评价提供了丰富的可供性(affordance)和新的发展契机。

基于此,本研究创新性地提出将视频游戏虚拟环境与空间能力测评相结合的研究设计和技术方案。其基本研究思路是:首先,在文献调研和因素分析的基础上,构建空间能力内在表征框架和评价指标体系。其次,使用多模态学习分析方法探索学习者空间认知行为表现特征模型。然后,使用以证据为中心的设计框架(evidence centered design, ECD)结合贝叶斯网络模型(Bayesian network, BN),构建整合“熟练度-证据-任务”三级变量网络的空间能力评价变量网络模型。之后,基于视频游戏引擎开发空间能力隐形测评工具,通过视频游戏虚拟环境中的互动任务将学习者内隐的思维过程显性化,并将测评算法嵌入程序,在测试程序后台自动收集学习者行为数据作为证据用以推断分析。最后,根据基于设计的研究(design-based research, DBR)方法,在实验室和真实课堂环境中开展实证研究,验证测评工具的有效性,并对测评工具循环迭代。根据以证据为中心的设计框架,本研究所选取的变量主要有三类:一是由空间能力理论与测评目标所确定的熟练度(proficiency)变量,如空间可视化、心理旋转、空间导航;二是由视频游戏测评工具设计特征所决定的任务(task)变量,如方向、选择操作、反应时、任务完成时间;三是由熟练度变量和任务变量之间因果关系所确定的证据(evidence)变量,如路线分布、漫游熵、选择正确率等(Almond et al., 2015)。在这三类变量中,任务变量是详细描述测评任务特征的参变量,证据变量是从学习者行为中直接测量得到的可观察变量,而熟练度变量通常是不能直接观察的潜变量,需要将证据变量代入证据识别规则和证据积累算法中进行推断和计算而得,以上三种变量共同构成了空间能力的“熟练度-证据-任务”三级变量网络。在此基础上,本研究将构建基于贝叶斯网络的空间能力的评价指标体系,并将其脚本部署在视频游戏后台,实现对空间能力的游戏化测评。

2 研究现状

2.1 基于游戏的测评方法

随着视频游戏技术和虚拟现实显示设备的不断发展和成熟,教育游戏与模拟仿真在教学实践中得到广泛应用,基于游戏的测评方法开始成为研究者关注的重点。基于游戏的测评(game-based

assessment, GBA)是指通过游戏或游戏化活动来对某个对象进行评价(Heinzen et al., 2015), 强调运用游戏设计原则来测量人们努力发挥到最佳水平时的表现。使用基于游戏的测评方法可以将视频游戏作为测试环境来对人的能力、人格等心理特征和行为表现进行量化评价(孙鑫 等, 2018)。根据测评的形式, 基于游戏的测评可以被分为游戏得分、外部测评、内嵌测评和严肃游戏分析四类(Kim & Ifenthaler, 2019)。其中, 游戏得分关注在游戏中实现目标或克服困难, 完成特定任务所用时间也是游戏得分指标; 外部测评是指在游戏环境之外开展的测试, 例如访谈、问卷、知识图谱或因果图; 内嵌测评则是游戏玩法的一部分, 并不会打断游戏体验, 通常是通过点击流、日志文件和信息轨迹中获取丰富的学习者行为数据; 严肃游戏分析和学习分析类似, 主要分析学习者个人特征(如社会背景信息、先验知识、技能和能力、兴趣等)和生成的游戏数据(如花费的时间、完成的目标或任务、导航模式、社交互动等) (Loh et al., 2015)。

近年来, 研究者开始使用基于游戏的测评方法来测量学习者的心理特征。其中, 最具代表性的是 Shute 等人提出的“隐形测评”, 即通过识别玩家在游戏内的行为指标证据来推断其潜在技能(Shute et al., 2015)。隐形测评强调使用不显眼的方法(如日志文件和眼动追踪)持续收集行为数据, 并检查学习者在游戏中的认知和非认知变量的发展, 模糊了游戏、学习和测评之间的界限(Shute & Ventura, 2013)。在此基础上, Shute 和 Rahimi (2021)在物理模拟游戏 Physics Playground 中使用以证据为中心的设计框架构建贝叶斯网络模型, 实现创造力的隐形测评, 并将其与游戏外部的创造力测评对比, 验证了测评工具的信效度。隐形测评的思想被广泛应用于基于游戏的测评研究。在认知能力方面, Quiroga 等(2015)选用 12 款商业视频游戏和 11 项智力能力测试来验证是否可以使用游戏来评价智力的一般因素(g)。因子分析结果表明, 视频游戏与智力测试能够测量相同的高阶潜在因子, 两者测试结果具有强相关性($r = 0.93$), 说明商业视频游戏能够可靠地测量个体的智力差异。Foroughi 等人使用视频游戏《传送门 2》内置的迷宫编辑器设计了一组含有 15 个游戏迷宫关卡的《传送门 2》测评量表, 用以评价玩家完成游

戏迷宫的能力水平, 发现其结果与瑞文高阶推理测试(RAPM)和 Bochumer 矩阵测试(BOMAT)所测得的流体智力变量具有强相关性($r = 0.78$) (Foroughi et al., 2016)。在行为动作方面, Roh 和 Lee (2014)使用体感游戏来评价注意缺陷多动障碍(ADHD)儿童与普通儿童的注意力, 发现 ADHD 儿童在漏击率、误击率、平均反应时和反应时标准差这四个指标上具有显著差异, 根据以上指标, 能够有针对性地开发三维体感游戏来提升儿童注意力。纽约大学 Wayne Mackey 等人开发了第一人称射击游戏训练平台 Aim Lab, 为玩家提供第一人称射击训练并提供实时反馈和诊断分析。研究者在后台收集超过 4000 万用户在射击任务中的行为表现数据, 以此来分析玩家在虚拟空间中长期运动学习的特点(Listman et al., 2021)。

基于游戏的测评还可以测量复杂问题解决和元认知等高阶思维过程。此类任务通常强调在真实情境中解决问题, 其测评方式始终与游戏紧密相连(张生 等, 2019)。例如, 经典模拟任务“裁缝店(Tailorshop)”就要求被试者扮演一家裁缝店的管理者, 通过控制工人数量、机器数量、调节原材料和广告等因素, 来尽可能实现更高产量、销售额和工人满意度, 这些指标就被作为复杂问题解决能力的测量指标(Danner et al., 2011)。香港中文大学开发的严肃游戏《农场狂想曲 2》构建了农场种植规划任务来支持复杂问题解决, 玩家需要选择一块土地, 然后根据当地气候选择投资的作物或动物, 并考虑温度、土壤、降水等环境因素, 通过施肥、浇水、喂食、洗澡等动作改变农产品参数, 获得尽可能多的收益(蒋宇 等, 2011)。Li 等人将视频游戏“推箱子(Sokoban)”改编为解谜任务, 并将“每个项目的计划时间与总时间之比”作为元认知计划的评价指标(Li et al., 2015)。

根据以上研究可以发现, 基于游戏的测评方法在内容开发、信息呈现、数据收集、可解释性、测评体验、应用推广等方面都具有显著优势: 首先, 视频游戏环境能够为学习者呈现丰富多样的信息刺激, 支持具身沉浸的交互体验, 提供适时有效的操作反馈, 从而将游戏内容与能力特征联系起来, 让游戏为个体提供思考和行动的线索(徐俊怡, 李中权, 2021)。其次, 通过游戏后台记录, 分析学习者在视频游戏环境中产生的、连续多样

的过程性行为表现数据,可以对学习者特征(知识、能力、态度等)进行测量和评价。这种测评工具可以嵌入游戏后台,实现沉浸式学习环境中的隐形测评(龚鑫等, 2023),不仅能够有效降低被试者的测验焦虑(Mavridis & Tsiatsos, 2017),也可以避免霍桑效应。第三,视频游戏可以通过画面和声音呈现学习者的行为表现,通过恰当的游戏机制设计,能够有效地将学习者内隐的认知过程通过交互行为可视化,直观呈现学生的学习策略和行为风格,并加强了数据的可解释性。第四,视频游戏的感官刺激和故事情节能够为学习者营造沉浸愉悦的使用体验,有助于提升参与动机和投入度(李一茗等, 2021)。最后,基于游戏的测评能够有效地降低测试参与门槛,并且可以通过在线平台大规模推广部署,易于形成大数据样本,为面向广大人群的认知能力测试和大数据研究奠定基础。

2.2 空间能力的游戏化测评研究

对于空间能力的评价始终是困扰研究者的难题。一方面,研究者对于空间能力的概念内涵和构成因素尚未形成共识,影响测评指标选取和量表开发。另一方面,空间能力涉及对心理表象的保持、操作和重构,而心理表象是个体对外部事物的内部再现,其认知加工过程难以通过客观有效的研究方法被直接测量。早期空间能力研究通常使用基于经验的评价,如言语报告、描画和再认等,但以上方法过于依赖被试者的自我内省和主观描述,缺少客观证据支持。随着上世纪心理测量学的发展,研究者开发了一系列经典空间能力测评量表,如心理旋转测试(mental rotation test, MRT) (Vandenberg & Kuse, 1978)、心理切割测试(mental cutting test, MCT) (College Entrance Examination Board, 1939)、普渡空间可视化测试(Purdue spatial visualization test: rotations, PSVT:R) (Bodner & Guay, 1997)、空间定向测试(spatial orientation test, SOT) (Friedman et al., 2020)、圣巴巴拉方向感测试(Santa Barbara sense of direction scale, SBSOD) (Hegarty et al., 2002)等。这些空间能力测评量表通常是从信息加工的角度出发,通过反应时和正确率指标来反映心理表象操作过程的快慢与难易程度,但测评指标的单一性也限制了对空间能力特征和内部机制的深入研究(刘鸣, 2004)。

近年来,计算机化测评技术的发展为空间能力测评领域带来了新的研究方向,研究者开始使用数字虚拟环境来训练和评价空间能力(郭守超, 2022; 尚俊杰等, 2022; Uttal et al., 2013)。具体来说:在基于对象的小尺度空间能力上,典型空间能力测试主要有心理旋转测试和空间可视化测试(包括心理切割、心理折叠和几何体组合任务等)。此类测试多数是经典纸笔量表的计算机化版本,与传统心理旋转测试纸质量表相比,基于3D虚拟对象的心理旋转能力测评能得到更高的评价准确率(Bartlett & Camba, 2023)。也有研究者使用3D建模工具来测评空间能力。例如,郭守超(2022)在3D CAD软件中自动采集学习者交互数据,系统分析了如何在三维虚拟环境中以数据驱动的测评方式来评价空间能力。在基于环境的大尺度空间能力上,典型的数字化测试包括虚拟Morris水迷宫测试(virtual Morris water maze, VMWM) (Thornberry et al., 2021)、八臂迷宫测试(4-on-8 Virtual Maze, 4/8 VM) (West et al., 2018),以及各类虚拟空间导航任务。此类测试通常包括学习阶段和测试阶段:在学习阶段,参与者要学习虚拟环境的空间位置,研究者记录其探索时间、探索距离等数据;在测试阶段,参与者进入虚拟空间完成测试任务,研究者记录其成功到达目标位置的正确率、所用时间、路线分布等指标,来对其空间导航和空间学习能力加以量化(张凤翔等, 2023)。例如,Waller (2005)要求被试者在虚拟环境中移动位置后重新指向物体位置,并识别建筑形状,以此来评价参与者基于视角转换的空间能力。Ventura等人开发的虚拟空间导航测试(virtual spatial navigation assessment, VSNA)根据完成导航和收集任务所用时间来评价空间能力(Ventura et al., 2013)。Weisberg等人开发的虚拟导航测试Virtual SILCton,要求被试者完成指向任务和模型构建任务,并自我报告导航过程(Weisberg et al., 2014)。West等(2018)构建基于VR场景的八臂迷宫导航任务,通过改变迷宫背景的环境线索,并观察被试者在改变前后完成任务所用时间,来判断其导航策略(反应策略或空间策略)。Malanchini等(2020)开发了虚拟导航测试Spatial Spy,在其中设置地图识别、路径记忆、地标导航、方向导航、扫视(scanning)和透视(perspective-taking)共6个任务,根据参与者能否在限定时间完成导航任务来

汇总计算其任务得分,并结合任务用时形成对空间导航能力的评价。以上任务通常是基于环境的路径探索,也有研究者通过量化个体行动轨迹来分析自由探索模式,并提出重访行为(revisit)、扩散模式(diffusion)、漫游熵(roaming entropy)及轨迹整合度(experienced integration)四个指标,重访行为代表探索的谨慎程度,扩散模式代表个体在区域内扩散的速率,漫游熵描述在给定时间内个体运动弥散程度,轨迹整合度描述个体探索环境中连接性较强部分的程度(Brunec et al., 2023; Gagnon et al., 2018)。以上指标可以用来描述和预测被试者空间能力水平。

基于视频游戏的空间能力测评研究是近几年出现的新趋势。此类研究的特点是使用商业游戏引擎或游戏编辑器开发游戏化测评工具,并且通常会开展验证性实验,将经典空间能力量表作为外部测试与新开发测评工具相对比,来验证其结果的有效性。例如, Peters 等人在 3D 益智游戏 Minecraft 的虚拟环境中开发了模式完形、心理旋转和空间构建任务,使用前者作为归纳推理的指标,后者作为空间能力的指标。结果显示,基于 Minecraft 的测试与常规心理测试(心理旋转测试 VMRT 与瑞文标准推理测试 RSPM)的结果潜变量高度相关($r = 0.72, p = 0.002$) (Peters et al., 2021)。Kim 等人基于 Unity 平台开发了可以在线评估空间推理能力的视频游戏 Shadowspect, 在小规模验证实验中收集了 44 名学生的游戏行为数据变量(包括“完成任务的时间、成功完成谜题的数量、失败后重返谜题的次数、已解决谜题的活动时间、失败尝试总数、操作形状的次数、Elo 平均能力分数”等 27 个变量指标)。之后,研究者使用随机森林回归模型计算了上述指标与两个外部测试(空间推理量表 SRI 和圣巴巴拉实体测试 SBST)得分的相关性,并确定模型中最重要的三个特征变量是“完成任务所用时间、持久性和能力得分”(Kim et al., 2023)。与此同时,也有研究者利用视频游戏能够吸引大量玩家主动参与的特点,使用移动社交软件实现面向大规模人群的空间能力游戏化测评。例如, Coutrot 等人(2019)将传统寻路任务和路径整合任务游戏化,开发了移动端游戏 Sea Hero Quest, 通过对游戏轨迹的分析来测量个体空间导航能力,并发现游戏测量结果与真实导航能力显著相关,具有良好生态效度。更进一步地,研究者

利用该款游戏在世界范围内收集到了 300 多万份玩家数据,通过大数据分析发现:性别、年龄、生活环境等因素对空间导航能力都存在影响(Coutrot et al., 2022)。Xu 等人(2023)基于微信平台开发了心理旋转测试小游戏《狼来了》,吸引公众参与心理旋转测试任务,获得了横跨多年龄段的大数据样本($N = 216, 713$)。基于此,研究者发现空间能力及其可塑性并非同步发展,心理旋转表现在 28 岁时达到峰值,而其可塑性早在 18 岁时就已达到峰值(Xu et al., 2023)。由此可见,基于游戏的空间能力测评具有诸多优势:一方面,视频游戏虚拟环境能够为学习者提供丰富多样的空间信息刺激,支持沉浸式、多模态、强反馈的人机交互体验。另一方面,视频游戏后台日志文件能为研究者提供细粒度的数据记录和分析,并且其轻松有趣的测评形式能够吸引广大用户群体主动参与,从而能以较低成本,在较短时间内获得大数据样本,实现面向空间能力的大规模测评研究。

2.3 空间能力的多模态测评研究

多模态(multimodal)概念最初由 Kress 和 Van Leeuwen (2001)提出,其核心定义为“在符号产品或事件中综合使用多种符号模态,以促进人类感官与外界环境的交互”。在教育领域,多模态技术的应用不断拓展,开始利用先进传感器技术和机器学习算法构建信息模型,来处理和分析复杂情境中的过程性学习数据(张琪 等, 2020)。近年来,随着多模态学习分析技术的进步,研究者得以将眼动追踪、生理信号、语音、视频和数字化日志等多源信息融合计算,全面理解学习者的学习和认知过程(田阳 等, 2019)。

在空间能力的测评研究中,多模态技术的应用日益广泛,涵盖了眼动追踪、脑电图(EEG)、功能性磁共振成像(fMRI)和功能性近红外光谱(fNIRS)等。这些技术从不同维度提供了生理和行为数据,为空间能力的全面评估提供了坚实的支持。眼动追踪技术在其中扮演了关键角色,提供了注视时间、注视点分布、扫视路径、瞳孔直径变化和注视次数等重要预测指标。这些指标不仅反映个体在执行空间任务时的注意力分配,还能揭示其认知策略(Fourtassi et al., 2017; Soares et al., 2022)。例如,在心理旋转任务中,瞳孔的显著扩张、较长的注视时间和复杂的扫视路径通常指示

个体在处理复杂空间信息时需要更多的认知资源。脑电图技术在空间能力测评中同样发挥着重要作用。脑电图能够记录大脑的电生理活动,通过分析事件相关电位(ERP)和事件相关去同步(ERD)现象,为评估空间能力提供了新的视角。特别是在心理旋转任务中, Beta 波段的 ERD 现象在顶叶区域尤为显著,反映了个体的空间认知能力和策略使用情况(陈鸿舟, 2015)。此外, EEG 信号的变化揭示了不同的认知机制,如前额叶和顶叶区域的活动差异,这可能与个体在任务执行过程中所采用的认知策略密切相关。此外,功能性磁共振成像和功能性近红外光谱技术的应用为研究大脑的空间表征和路径选择提供了强有力的工具。研究表明,海马区与顶叶、前额叶的协同作用对于实现有效的空间导航至关重要(Kong et al., 2017)。通过整合 fMRI 数据与眼动追踪数据,研究者能够深入分析个体在导航任务中的实时脑活动和行为表现,为揭示空间导航的认知过程提供了更为丰富和立体的信息(Ardila, 2018)。然而,每一种单独的技术在实际研究中都存在一定的优势和局限性。例如, EEG 技术虽然具有高时间分辨率,能实时反映大脑活动,但空间分辨率较低,脑功能溯源分析不够精确。而眼动数据虽然能反映当前的注视过程,但难以揭示大脑的内在加工机制。因此,将 fMRI 与眼动追踪和 EEG 相结合的研究手段越来越受到神经认知领域研究人员的重视。多模态技术的融合为空间能力的全面评估提供了新的可能性,通过整合眼动数据、脑电信号和脑影像数据,可以更深入地了解个体在空间任务中的认知过程。

多模态学习分析更侧重于收集学习者在真实物理环境中的行为数据和生理信号,以补充或验证虚拟环境中收集的数据(汪维富, 毛美娟, 2021)。显然,多模态学习分析可以整合学习者在解决空间任务时的多模态、互动的数字痕迹,以此为基础,实现基于多模态数据的空间能力动态评估和反馈。此外,结合人工智能方法可以进一步理解、解释和推理学习者的空间能力。例如,在心理旋转任务中,结合眼动数据和脑电信号,可以揭示个体在不同策略下的认知加工模式(Xue et al., 2017)。结合多模态数据能够提供更准确的空间能力预测模型,例如通过机器学习算法整合不同模态的数据,能够提高预测的准确性和稳健性

(Fourtassi et al., 2017)。因此,将学习者的生理数据(如眼动、脑电、表情)与虚拟行为数据(如数字化日志、录屏)结合,可通过多模态学习分析方法揭示背后的认知机制,为基于视频游戏的空间能力评估工具提供理论支持。

3 研究构想

本研究旨在准确、有效、大规模地测评空间能力,将采用多模态学习分析、贝叶斯网络计算建模、随机对照实验与因素分析、基于设计的研究等跨学科研究方法,结合视频游戏所收集的学习者空间能力行为表现数据样本,研究基于视频游戏的空间能力测评技术方案,并开发面向空间能力的游戏化测评工具,研究框架如图1所示。

首先,研究1通过文献调研与因素分析,确定空间能力的概念、分类、表征形式和经典测评指标,确定空间能力的熟练度模型;研究2利用多模态学习分析方法探明学习者在三维虚拟空间中解决空间问题时的行为表现特征,计算构建空间能力行为表现模型,以此确定适用于测评的可观察变量指标和空间能力任务模型;研究3通过文献调研和验证性对比实验,确定视频游戏影响空间能力的关键因素(如界面维度和任务类型),明确视频游戏的人机交互特征,并使用游戏引擎开发基于视频游戏的测评工具,完成空间能力的呈现模型;研究4将在前三者的基础上,采用计算建模与机器学习方法,结合以证据为中心的设计框架(ECD)与贝叶斯网络模型(BN),设计并开发能够推断和预测空间能力的测评算法,并将其嵌入游戏后台程序之中,实现对空间能力的隐形测评;研究5将空间能力测评工具分别在实验室和中小学课堂中应用,通过随机对照实验与真实情境下的教学实践,验证空间能力测评工具的有效性,并探索能够在实验室和真实课堂中有效应用的测评模式。

3.1 研究1: 空间能力内在表征框架与评价指标体系构建

本部分将重点关注空间能力的概念、分类、表征形式和经典测评指标,以确定空间能力内在表征框架和变量网络,并构建对应的评价指标体系。具体包括:调研空间能力测评相关的理论和实证研究,明确空间能力的核心概念、影响因素、表征形式和经典测评量表。首先,根据空间能力

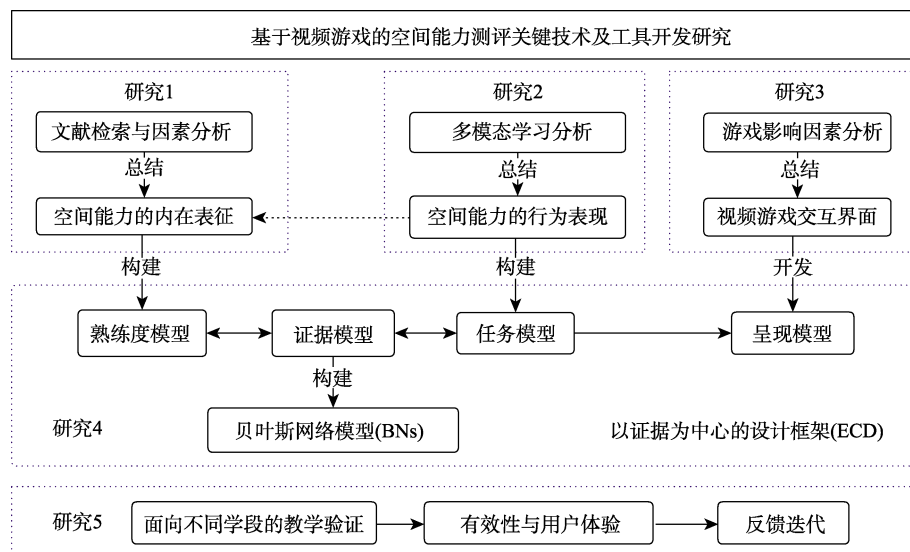


图 1 本研究整体框架

经典理论,初步确定空间能力的分类框架和维度,如认知过程、空间尺度、空间维度、动态静态等。之后,在先前研究基础上,收集空间能力经典测试量表,并在实验室环境中开展验证性实验,通过因素分析,探索各类指标变量之间的相关关系和从属结构,以此构建变量关系网络和内在表征框架,形成空间能力的熟练度模型。基于空间能力变量关系网络和表征框架,梳理经典空间能力评价量表和评价任务,并分析不同空间能力测评的认知加工过程,明确不同空间能力测评指标对空间能力因子的预测关系。总的来说,研究 1 主要是基于经典心理测量学范式,通过文献调研从已有研究中明确空间能力基本概念与分类,总结并归纳各类空间能力的心理表征、典型任务和关键测评指标。然后开展验证实验,通过因素分析,确定各类测评指标之间的相关性和结构关系,初步形成空间能力变量关系网络,为研究 4 空间能力贝叶斯网络的构建奠定基础。根据空间能力分类框架,可以被计算机化测评的各类空间能力测评任务及其潜在测评指标如表 1 所示。

3.2 研究 2: 基于多模态学习分析的空间能力行为表现模型

本部分将采用多模态学习分析方法确定学习者在三维虚拟空间中解决空间问题时的行为表现特征,计算构建空间能力行为表现模型,以此确定适用于测评的可观测变量,进一步明确空间能

力的任务模型。具体包括: 1) 使用视频游戏编辑工具或游戏引擎设计空间能力测试任务集。该测试任务集围绕空间能力的可观测变量构建,通常是由一系列由易到难的游戏任务关卡组成。例如,在三维虚拟环境中构建一组迷宫任务,在其中设计定向导航任务和地标导航任务来收集学习者行为数据的可观察变量(如任务用时、行动轨迹、无效探索时间、完成任务数量、游戏内得分等指标)。通过改变导航任务中导航距离、地标线索、信息提示等要素,来考察被试者在不同条件下的导航过程。之后,根据研究者确立的证据规则,对以上可观察变量指标进行证据识别和证据汇总,作为评价和预测学习者空间能力水平的证据变量。2) 使用眼动追踪和脑电图(EEG)采集被试者实时生理信号开展多模态学习分析。为获得学习者在真实物理环境中的行为数据和生理信号,以补充或验证虚拟环境中收集的数据,在上述研究过程中,使用多模态学习分析方法将学习者的多模态生理数据(如眼动、脑电、表情等)与游戏行为数据(如鼠标键盘操作数据流、游戏动作录像、游戏录屏画面等)相结合。整合学习者在解决空间问题过程中产生的多模态、互动的数字痕迹,以此实现基于多模态数据的空间能力动态评估和反馈。例如,在时间轴上将多模态数据同步,根据学习者在完成迷宫任务时的眼动指标(首次注视时间,注视总时间、注视点分布、注视次数、扫视路径和瞳孔

表 1 空间能力测评任务及其潜在测评指标

空间能力一级维度	空间能力二级维度	空间能力测评任务	潜在测评指标
对象操纵	心理旋转	心理旋转测试(VMRT) (Xu et al., 2023) 圣巴巴拉固体测试(SBST) (Cohen & Hegarty, 2012) 空间构建任务(Peters et al., 2021)	反应时
	心理折叠		正确率
对象操纵	心理切割		几何体差异度
	几何体组合		有效运动距离
	空间可视化		距焦点物体距离 垂直高度可变性
扫视透视	扫视 视角选取 空间定向	扫视识别任务(Malanchini et al., 2020)	反应时 正确率 角度误差
		多面体投影识别任务(Waller, 2005)	
		圣巴巴拉方向感测试(SBSOD) (Hegarty et al., 2002)	
		空间定向测试(SOT) (Friedman, et al., 2020)	
		皮亚杰三山测试(McLaren-Gradinaru et al., 2020)	
空间导航	地标导航 方向导航 路径记忆 地图识别	空间布局测试(SCT) (McLaren-Gradinaru et al., 2020)	反应时 正确率 移动距离 路线分布 重访行为 扩散模式 漫游熵 轨迹整合度
		虚拟 Morris 水迷宫(VWM)	
		虚拟空间导航测试(VSNA) (Ventura et al., 2013)	
		八臂迷宫任务(4/8 Maze) (West et al., 2018)	
		传送门测试(Portal test) (Foroughi et al., 2016)	

直径变化等)、脑电指标(α 波、 θ 波、N2、P2、P3波幅,顶叶区域 Beta 波段的 ERD 现象)分析学习者解决空间问题时的生理表现特征,基于此提出多模态数据驱动的空间能力模型。3)通过与空间能力外部测试对比验证本研究证据变量的有效性。使用经典空间能力测试(如 Morris 水迷宫测试、心理旋转测试)对同一批被试者展开外部测试,将测试结果作回归分析,验证本研究空间能力测评任务的有效性。

3.3 研究 3: 视频游戏影响因素分析与游戏化测评工具开发

本部分通过文献调研和验证性对比实验,确定视频游戏影响空间能力的关键因素,明确视频游戏的人机交互特征,并使用视频游戏引擎开发基于游戏的测评工具,完成空间能力的呈现模型(presentation model);具体来说,1)在文献调研的基础上,开展增值研究与认知后果研究,探讨“哪些游戏功能和特征能够影响空间能力”以及“玩某种游戏是否会影响空间能力”,以此确定能够影响各类空间能力的视频游戏类型(Mayer, 2014)。石祝和尚俊杰(2024)分析了各类视频游戏的媒介特征,并构建了视频游戏影响空间能力的理论机制模型。研究显示,视频游戏影响学习者空间能

力的关键要素是游戏界面空间维度(2D/3D)和游戏玩法要求的操作速度(actions per minute, APM),游戏界面要能够提供与测评任务空间维度相一致的空间刺激,而游戏玩法要能够激发学习者与测评任务相对应的认知加工过程。在此视角下,三维益智游戏(3D puzzle games)兼具三维虚拟空间场景和较慢的游戏节奏,既能提供与三维心理旋转、空间可视化、空间导航等测试任务相匹配的空间刺激,又能鼓励学习者使用空间认知策略探索环境,是最有可能影响空间能力的游戏类型。2)使用视频游戏引擎(如 Unreal 5、Unity3D)设计并开发测评空间能力的三维益智游戏。此游戏将在三维虚拟空间场景中构建各类空间能力测试关卡,玩家以第一人称视角控制虚拟化身(avatar)在其中探索并完成各类空间能力测试任务。测试任务根据表 1 所述,将经典测试任务改编成为游戏关卡,针对对象操作、扫视透视、空间导航这三种空间能力分别设计一系列由易到难的游戏关卡作为空间能力的测评单元。最后,设计游戏背景故事将这些关卡整合为一个完整的视频游戏,实现空间能力的呈现模型。该呈现模型确定了视频游戏的交互界面,并限定了玩家操作行为的边界,以此可以进一步明确证据模型应当选取哪些行为

表现和后台数据作为判断空间能力水平的可观察变量,并为测评算法开发奠定基础。

3.4 研究 4: 基于贝叶斯网络模型的空间能力测评算法开发

研究 4 将采用计算建模与机器学习方法,结合以证据为中心的设计框架与贝叶斯网络模型,构建空间能力内在心理表征(熟练度模型)与外在行为表现(任务模型)变量之间的证据关系(证据模型),设计并开发能够推断和预测空间能力的测评算法,并将其嵌入视频游戏后台程序之中,实现对空间能力的隐形测评。具体来说,第一步是构建空间能力测评算法,根据 ECD 框架来组织变量关系,该框架的优点在于能够在复杂表现任务情况下对复杂能力展开测量。因此适用于空间能力的游戏化测评(Choi & Mislevy, 2022)。ECD 框架的三个核心变量模型分别是熟练度模型(也称学生模型)、证据模型和任务模型,三个模型协同运作,共同构成“熟练度-证据-任务”三级变量网络,并最终以呈现模型的形式交付使用(Almond et al., 2015)。在本研究中,如表 1 所示,空间能力内在表征的一、二级分类维度对应熟练度变量,空间能力外在行为表现对应证据变量(如反应时、正确率等),而任务变量体现在测评任务的设计之中(如任务难度、虚拟代理数量等)。

本研究选取贝叶斯网络进行分析和数据建模,不仅是因为其适用于 ECD 框架,而且是因为其在基于游戏的测评中具有诸多优势(de Klerk et al., 2015)。一是能直观展现变量关系,提供图形化的熟练度模型,便于整合现有理论和专家知识;二是作为概率模型,可从数据中学习并随数据增加不断优化;三是能从游戏环境中获取学习者表现数据,提供多粒度实时分析;四是有可靠软件支持,便于构建复杂大型网络,降低分析难度(Almond et al., 2015)。此外,贝叶斯网络可以使用离散变量快速评价,适合嵌入到游戏引擎之中。其构建过程如下:1)根据领域分析(即研究 1)所确定的空间能力维度和变量关系构建贝叶斯网络模型原型,将空间能力分类变量作为熟练度变量,搭建贝叶斯网络的一级和二级节点;其次,在潜在评价指标中(见表 1),根据研究 2 和研究 3,确定适用于空间能力测评的可观察变量(即证据变量)和任务变量,并将二者共同作为贝叶斯网络最低层级的子节点;第三,设计证据规则,按照证

据规则将证据变量识别转化为中介变量,并进一步汇总计算,从而推断空间能力的熟练度变量,从而使贝叶斯网络一、二级节点和底层节点依据变量关系连接起来。在形成贝叶斯网络基本结构之后,需要根据现有研究证据,定义每个节点的先验概率分布(通常使用正态分布),使用研究 2、研究 3 中学习者空间探索的实验数据对其进行预训练,确定变量节点间的概率分布,检查并完善网络。2)将所有通过计算机自动化收集的交互数据按照 Experience API (简称 xAPI)数据规范整理重构,将行为变量以“用户-对象-操作”的格式加以汇总,以实现多源数据融合的数据建模方法(武法提,黄石华,2020)。之后,将贝叶斯网络嵌入到游戏后台程序之中,每当玩家完成一个游戏关卡时,系统将自动调取日志文件,对空间能力贝叶斯网络的各个子节点(证据变量和任务变量)生成新的证据,而父节点(即中介变量或熟练度变量)则会根据证据识别和证据积累过程自动更新,从而实现对空间能力整体和各个子维度能力水平的概率估计。该测评工具将能够自动记录学习者的行为数据(如在游戏中的位置、观察方向和按键操作流),以此来实时追踪其互动行为,并能将反映空间能力水平的可观测变量数据作为证据输入到贝叶斯网络中,从而更新空间能力水平的概率预测。最后,通过用户界面将评价结果可视化反馈给教育者和学习者。

3.5 研究 5: 空间能力测评工具的验证研究与教学实践

本部分聚焦空间能力游戏化测评的验证与优化,旨在通过验证性实验与教学实践,验证本研究开发的空間能力测评工具的有效性。首先,采用心理测量学量表的经典方法,在实验室环境中招募参与者(以在校大学生为主)开展随机对照实验,在填写基本背景信息和所学专业后,分别完成本研究开发的空間能力游戏化测评和空間能力经典测评。为控制测评媒介带来的无关变量,也便于收集和处理数据,本研究使用计算机化的空間能力经典测评作为外部测试,包括心理旋转测试(VMRT)、空間定向测试(SOT)、圣巴巴拉固体测试(SBST)以及虚拟水迷宫测试(VWM)等。通过一致性检验、随机森林回归、多元线性回归或结构方程模型,对比不同工具测评结果之间的相关性,确定空間能力游戏化测评工具的有效性。将

实验结果数据纳入贝叶斯网络模型,来进一步训练和更新空间能力测评系统的参数。

之后,研究者将使用基于设计的研究方法,在真实教育情境中与中小学教师长期合作,对游戏化测评工具开展教学实践,通过设计、开发、反复实施的迭代循环,确保其能够准确、有效且高效地测评学习者的空间能力(梁林梅等, 2022)。本部分的目标是通过广泛的教学实践来验证本研究开发的空間能力评估工具在各教育阶段的有效性。具体来说,将依据本课题组在游戏化学习方面的研究经验,开发用户体验调查问卷和技术接受度问卷展开调研,并辅以一一对一访谈和焦点小组访谈,收集教师和学生对测评工具的反馈建议,来对测评工具不断迭代和优化,并在此基础上探索在真实课堂环境中使用空间能力游戏化测评工具的教学模式。

4 理论建构与创新

空间能力是与语言能力、数字能力并列的人类基本认知能力,被称为“所有思考的根基”(周加仙,蔡永华, 2013)。空间能力不仅影响学习者在数学、科学等学科的学业表现,而且能够预测他们未来在STEM领域的职业发展(Wai et al., 2009)。因此,准确、有效且高效地测评空间能力,不仅对心理学、认知神经科学、脑科学等研究领域具有重要的科学意义,而且对于科技创新、航空航海等专业领域的人才选拔和培养具有重要的教育意义和应用价值。由于空间能力具有复杂性、多维度、内隐性的特点,一直以来,空间能力的测量和评价都是困扰研究者和教育者的难题。已有研究通常是将空间能力经典纸笔量表数字化,普遍存在脱离真实情境、数据维度单一、信息量较小、可解释性与可迁移性差等问题,难以实现对空间能力的整体理解(Uttal et al., 2013)。当前已经有研究使用三维虚拟空间来测评空间导航能力,丰富和拓展了空间能力的测评指标(Malanchini et al., 2020; Ventura et al., 2013; Weisberg et al., 2014; West et al., 2018)。更进一步地,研究者也开始尝试基于视频游戏开发空间能力测评工具,这不仅拓展了空间能力的过程性指标(Kim et al., 2023; Peters et al., 2021),而且能收集到传统测评工具难以获得的大数据样本(Coutrot et al., 2019; Xu et al., 2023)。

本研究在前人研究基础上进一步发展了基于视频游戏的空间能力测评方法。相比于先前的游戏化测评研究多聚焦于对游戏后台数据的事后分析,本研究的创新在于将游戏化测评与多模态学习分析深度融合,实现了对空间能力的过程性评价和整体性理解。首先,在自主开发的视频游戏测评工具中,创造性地将以证据为中心的设计框架和贝叶斯网络相结合,用于识别和汇总能够推断学习者空间能力水平的多模态行为数据,能够充分利用已有研究和专家背景知识,并增强了测评模型的可识别性与可解释性。其次,创新性地提出多源数据驱动的空间能力测评范式,构建了整合学习者行为数据(观察数据)、游戏后台数据(交互数据),以及游戏过程中的生理信号数据(情境数据)的多源数据采集机制(郭守超, 2022),并将以上数据在时间轴上对准,根据证据规则纳入到推断系统中,更新对空间能力的概率预测,实现空间能力的自动化测评。再者,本研究深入分析视频游戏影响空间能力的认知机制,使用三维益智游戏(3D puzzle games)作为训练和测评空间能力的虚拟环境,并在视频游戏中使用xAPI规范自动采集学习者交互数据,实现空间能力的隐形测评,可以有效避免测评对学习过程的影响,以及测试焦虑、霍桑效应等人因误差,兼顾内在效度与生态效度,为未来游戏化空间能力测评工具的推广应用奠定技术基础。最后,本研究旨在实现对空间能力的整体理解,力求在基于游戏的测评中覆盖空间能力不同尺度、不同认知过程的特征变量,通过测评结果进一步丰富和拓展了空间能力的概念和内涵,为空间能力理论研究提供工具支撑。

本研究使用以证据为中心的设计范式构建了一套独创性的测评工具,旨在深化对空间能力的认知。在理论研究基础上,借助前沿的多模态分析技术,在游戏互动的场景下,同步采集脑电图、眼球追踪等多元生理信号,同时整合游戏后台的操作日志及行为表现指标。依据“熟练度-证据-任务”的框架结构,将空间能力各级各类变量整理汇总到贝叶斯网络模型之中,构建空间能力的隐形测评,实现对空间能力的整体理解。本研究所创设的游戏化测评工具,不仅有效规避了传统心理测试中常见的局限——如缺乏过程数据、易受社会期望偏差影响等弊端,更巧妙地利用了视

频游戏作为大众媒体的强大吸引力,具备低成本、高覆盖的优势。借助社交网络与互联网平台的便捷传播特性,游戏化测评能够激发广大玩家的参与热情,进而收集到海量数据,其样本规模可达万、十万甚至百万级。如此庞大的数据样本为认知能力的大规模群体性研究提供了前所未有的机遇。综上,本研究设计的基于视频游戏的空间能力测评,有望攻克长久以来空间能力难以量化与评估的难题,对于大范围人才选拔与培育工作具有深远意义,并能为空间能力研究领域带来新的突破。

参考文献

- 陈鸿舟. (2015). 基于脑电的不同策略下心理旋转的认知机制研究 [硕士学位论文]. 上海交通大学.
- 龚鑫, 许洁, 乔爱玲. (2023). 基于沉浸式学习环境的隐性评估: 机理、框架与应用. *电化教育研究*, (12), 64-72.
- 郭守超. (2022). 数据驱动的空间能力评价研究 [博士学位论文]. 华东师范大学, 上海.
- 蒋宇, 尚俊杰, 庄绍勇. (2011). 游戏化探究学习模式的设计与应用研究. *中国电化教育*, (5), 84-91.
- 李洪玉, 林崇德. (2005). 中学生空间认知能力结构的研究. *心理科学*, 28(2), 269-271.
- 李一茗, 杨上琦, 黎坚. (2021). 基于游戏的评估: 特殊儿童心理评估的新方向. *中国特殊教育*, (12), 90-96.
- 梁林梅, 蔡建东, 耿倩倩. (2022). 学习科学研究与教育实践变革: 研究方法论的创新和发展. *电化教育研究*, 43(1), 39-45+62.
- 刘鸣. (2004). 表象研究方法论. *心理科学*, 27(2), 258-260.
- 尚俊杰, 曾嘉灵, 周均奕. (2022). 学习科学视角下的数学空间游戏设计与应用研究. *电化教育研究*, 43(7), 63-72.
- 石祝. (2019). 视频游戏对中学生空间能力和问题解决能力的影响: 基于视频游戏《传送门2》的实证研究 [硕士学位论文]. 北京师范大学.
- 石祝, 尚俊杰. (2024). 视频游戏对空间能力的影响与作用机制. *中国电化教育*, (5), 32-44+113.
- 孙鑫, 黎坚, 符植煜. (2018). 利用游戏 log-file 预测学生推理能力和数学成绩——机器学习的应用. *心理学报*, 50(7), 761-770.
- 田阳, 陈鹏, 黄荣怀, 曾海军. (2019). 面向混合学习的多模态交互分析机制及优化策略. *电化教育研究*, 40(9), 67-74.
- 汪维富, 毛美娟. (2021). 多模态学习分析: 理解与评价真实学习的新路向. *电化教育研究*, (2), 25-32.
- 武法提, 黄石华. (2020). 基于多源数据融合的共享教育数据模型研究. *电化教育研究*, 41(5), 59-65+103.
- 徐俊怡, 李中权. (2021). 基于游戏的心理测评. *心理科学进展*, 29(3), 394-403.
- 晏碧华. (2021). 飞行员空间能力: 认知计算的可塑性与渗透性. 北京: 科学出版社.
- 张凤翔, 陈美璇, 蒲艺, 孔祥祯. (2023). 空间导航能力个体差异的多层次形成机制. *心理科学进展*, 31(9), 1642-1664.
- 张琪, 李福华, 孙基男. (2020). 多模态学习分析: 走向计算教育时代的学习分析学. *中国电化教育*, (9), 7-14+39.
- 张生, 任岩, 骆方. (2019). 学生高阶思维能力的评价: 复杂问题解决的测量述评. *中国特殊教育*, (10), 90-96.
- 周加仙, 蔡永华. (2013). 空间思维能力的认知机制及其对STEM教学的启示——与美国 Nora S. Newcombe 院士的对话. *全球教育展望*, 42(2), 3-11.
- Açıkgül, K., Şad, S. N., & Altay, B. (2023). Spatial ability test for university students: Development, validity and reliability studies. *International Journal of Assessment Tools in Education*, 10(1), 76-97.
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). *The conceptual assessment framework*. In *Bayesian networks in educational assessment*. Springer.
- Ardila, A. (2018). *Historical development of human cognition: A cultural-historical neuropsychological perspective* (Vol. 3). Springer.
- Bartlett, K. A., & Camba, J. D. (2023). Gender differences in spatial ability: A critical review. *Educational Psychology Review*, 35(1), 8.
- Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, 2, 1-17.
- Brunec, I. K., Nantais, M. M., Sutton, J. E., Epstein, R. A., & Newcombe, N. S. (2023). Exploration patterns shape cognitive map learning. *Cognition*, 233, 105360.
- Choi, Y., & Mislevy, R. J. (2022). Evidence centered design framework and dynamic Bayesian network for modeling learning progression in online assessment system. *Frontiers in Psychology*, 13, 742956.
- Cohen, C. A., & Hegarty, M. (2012). Inferring cross sections of 3D objects: A new spatial thinking test. *Learning and Individual Differences*, 22(6), 868-874.
- College Entrance Examination Board. (1939). *CEEB special aptitude test in spatial relations*. Author.
- Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., ... Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904), 104-110.
- Coutrot, A., Schmidt, S., Coutrot, L., Pittman, J., Hong, L., Wiener, J. M., ... Spiers, H. J. (2019). Virtual navigation tested on a mobile APP is predictive of real-world wayfinding navigation performance. *Public Library of Science ONE*, 14(3), e0213272.
- Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring performance in dynamic decision making: Reliability and validity of the tailorshop simulation. *Journal of Individual Differences*, 32(4), 225-233.
- de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015).

- Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & Education*, (85), 23–34.
- Foroughi, C. K., Serraino, C., Parasuraman, R., & Boehm-Davis, D. A. (2016). Can we create a measure of fluid intelligence using puzzle creator within Portal 2? *Intelligence*, 56, 58–64.
- Fourtassi, M., Rode, G., & Pisella, L. (2017). Using eye movements to explore mental representations of space. *Annals of Physical and Rehabilitation Medicine*, 60(3), 160–163.
- Friedman, A., Kohler, B., Gunalp, P., Boone, A. P., & Hegarty, M. (2020). A computerized spatial orientation test. *Behavior Research Methods*, 52, 799–812.
- Gagnon, K. T., Thomas, B. J., Munion, A., Creem-Regehr, S. H., Cashdan, E. A., & Stefanucci, J. K. (2018). Not all those who wander are lost: Spatial exploration patterns and their relationship to gender and spatial memory. *Cognition*, 180, 108–117.
- Geer, E. A., Quinn, J. M., & Ganley, C. M. (2019). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology*, 55(3), 637–652.
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5), 425–447.
- Heinzen, T. E., Landrum, R. E., Gurung, R. A. R., & Dunn, D. S. (2015). Game-based assessment: The mash-up we've been waiting for. In T. Reiners & L. C. Wood (Eds.), *Gamification in education and business* (pp. 201–217). Springer International Publishing.
- Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology*, 88(4), 675–697.
- Hong, X., & Liu, Q. (2022). Assessing young children's national identity through human-computer interaction: A game-based assessment task. *Frontiers in Psychology*, 13, 956570.
- Hunt, E., & Pellegrino, J. (1985). Using interactive computing to expand intelligence testing: A critique and prospectus. *Intelligence*, 9(3), 207–236.
- Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, 24(9), 1831–1836.
- Kim, Y. J., & Ifenthaler, D. (2019). Game-based assessment: The past ten years and moving forward. In: Ifenthaler, D., Kim, Y. J. (Eds.), *Game-based assessment revisited*. Advances in Game-Based Learning. Springer, Cham. https://doi.org/10.1007/978-3-030-15569-8_1
- Kim, Y. J., Knowles, M. A., Scianna, J., Lin, G., & Ruipérez-Valiente, J. A. (2023). Learning analytics application to examine validity and generalizability of game-based assessment for spatial reasoning. *British Journal of Educational Technology*, 54(1), 355–372.
- Kim, Y. J., & Shute, V. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, 87, 340–356.
- Kong, X. Z., Wang, X., Pu, Y., Huang, L., Hao, X., Zhen, Z., & Liu, J. (2017). Human navigation network: The intrinsic functional organization and behavioral relevance. *Brain Structure and Function*, 222(2), 749–764.
- Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication* (pp. 20–22). Arnold Publishers.
- Li, J., Zhang, B., Du, H., Zhu, Z., & Li, Y. M. (2015). Metacognitive planning: Development and validation of an online measure. *Psychological Assessment*, 27(1), 260–271.
- Listman, J. B., Tsay, J. S., Kim, H. E., Mackey, W. E., & Heeger, D. J. (2021). Long-term motor learning in the “wild” with high volume video game data. *Frontiers in Human Neuroscience*, 15, 777779.
- Liu, S., Wei, W., Chen, Y., Hugo, P., & Zhao, J. (2021). Visual-spatial ability predicts academic achievement through arithmetic and reading abilities. *Frontiers in Psychology*, 11, 591308.
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious games analytics: Theoretical framework. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious games analytics. Methodologies for performance measurement, assessment, and improvement* (pp. 3–29). Springer.
- Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, 49(4), 344–351.
- Malanchini, M., Rimfeld, K., Shakeshaft, N. G., McMillan, A., Schofield, K. L., Rodic, M., ... Plomin, R. (2020). Evidence for a unitary structure of spatial cognition beyond general intelligence. *NPJ Science of Learning*, 5, 9.
- Martin-Dorta, N., Saorín, J. L., & Contero, M. (2008). Development of a fast remedial course to improve the spatial abilities of engineering students. *Journal of Engineering Education*, 97(4), 505–513.
- Mavridis, A., & Tsiatsos, T. (2017). Game-based assessment: Investigating the impact on test anxiety and exam performance. *Journal of Computer Assisted Learning*, 33(2), 137–150.
- Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. Boston Review.
- McLaren-Gradinaru, M., Burles, F., Dhillon, I., David, A. L., Umiltà, A., Hannah, J., Dolhan, K., & Iaria, G. (2020). A novel training program to improve human spatial

- orientation: Preliminary findings. *Frontiers in Human Neuroscience*, 14, 5.
- Owen, V. E., & Baker, R. S. (2020). Fueling prediction of player decisions: Foundations of feature engineering for optimized behavior modeling in serious games. *Technology, Knowledge and Learning*, 25(2), 225–250.
- Peters, H., Kyngdon, A., & Stillwell, D. (2021). Construction and validation of a game-based intelligence assessment in Minecraft. *Computers in Human Behavior*, 119(5), 106701.
- Quiroga, M. Á., Escorial, S., Román, F. J., Morillo, D., Jarabo, A., Privado, J., ... Colom, R. (2015). Can we reliably measure the general factor of intelligence (g) through commercial video games? Yes, we can! *Intelligence*, 53, 1–7.
- Roh, C. H., & Lee, W. B. (2014). A study of the attention measurement variables of a serious game as a treatment for ADHD. *Wireless Personal Communications*, 79, 2485–2498.
- Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, 116, 1–13.
- Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in video games: Stealth assessment*. Cambridge, MA: The MIT Press.
- Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Computers & Education*, 80, 58–67.
- Soares Jr, R. da S., Oku, A. Y. A., Barreto, C. S. F., & Sato, J. R. (2022). Applying functional near-infrared spectroscopy and eye-tracking in a naturalistic educational environment to investigate physiological aspects that underlie the cognitive effort of children during mental rotation tests. *Frontiers in Human Neuroscience*, 16, 889806.
- Sorby, S. A., & Panther, G. C. (2020). Is the key to better PISA math scores improving spatial skills? *Mathematics Education Research Journal*, 32(2), 213–233.
- Spiers, H. J., Coutrot, A., & Hornberger, M. (2023). Explaining world-wide variation in navigation ability from millions of people: Citizen science project Sea Hero Quest. *Topics in Cognitive Science*, 15(1), 120–138.
- Thornberry, C., Cimadevilla, J. M., & Commins, S. (2021). Virtual Morris Water Maze: Opportunities and challenges. *Reviews in the Neurosciences*, 32(8), 887–903.
- Uttal, D. H., & Cohen, C. A. (2012). Chapter Four - Spatial thinking and STEM education: When, why, and how? *Psychology of Learning and Motivation*, 57, 147–181.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2), 599–604.
- Ventura, M., Shute, V., Wright, T., & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology*, 4, 852.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835.
- Waller, D. (2005). The WALKABOUT: Using virtual environments to assess large-scale spatial abilities. *Computers in Human Behavior*, 21(2), 243–253.
- Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3), 669–682.
- West, G. L., Konishi, K., Diarra, M., Benady-Chorney, J., Drisdelle, B. L., Dahmani, L., ... Bohbot, V. D. (2018). Impact of video games on plasticity of the hippocampus. *Molecular Psychiatry*, 23(7), 1566–1574.
- Xie, F., Zhang, L., Chen, X., & Xin, Z. (2020). Is spatial ability related to mathematical ability: A meta-analysis. *Educational Psychology Review*, 32, 113–155.
- Xu, S., Song, Y., & Liu, J. (2023). The development of spatial cognition and its malleability assessed in mass population via a mobile game. *Psychological Science*, 34(3), 345–357.
- Xue, J., Li, C., Quan, C., Lu, Y., Yue, J., & Zhang, C. (2017). Uncovering the cognitive processes underlying mental rotation: An eye-movement study. *Scientific Reports*, 7(1), 10076.

Video game-based assessment of spatial ability

SHANG Junjie, SHI Zhu, SHEN Kejie

(Lab of Learning Sciences, Graduate School of Education, Peking University, Beijing 100871, China)

Abstract: Spatial ability refers to the ability of individuals to recognize, encode, store, represent, decompose, combine and abstract objects or spatial figures in their minds, which is the cognitive foundation for understanding one's environment and solving problems. Building an accurate, convenient and effective assessment system of spatial ability is of great significance to the enhancement of STEM education and the quality of talent cultivation. Due to the complex, multi-dimensional and implicit nature of spatial ability, it is difficult to evaluate spatial ability via computer-based assessments. This study aims to accurately, effectively, and massively evaluate spatial ability by using multimodal learning analytics methods to explore the characteristic behavioral expressions of learners' spatial cognition, and by developing key technologies and tools for spatial ability stealth assessment based on video game environments. The specific contents include: 1) Construct a framework for the intrinsic representation of spatial ability and an evaluation index system; 2) Constructing a learner spatial ability behavior performance model based on multimodal learning analysis; 3) Explore the key factors that influence spatial ability in video games, and use game engines to develop game-based assessment tools; 4) Use evidence-centered design frameworks and Bayesian network models to develop and deploy assessment algorithms capable of inferring and predicting spatial abilities; 5) Conduct empirical research in laboratory and real classroom settings to verify the effectiveness of evaluation tools. The research findings will contribute to a better understanding of human spatial cognition processes and behavioral performance, expand and enrich theories related to spatial abilities, and provide key technical support for large-scale digital assessment.

Keywords: game-based assessment, spatial ability, multimodal learning analysis, game-based learning, stealth assessment