

从招聘选拔到工作表现：认知能力的差异化预测效应*

邹雨含¹ 管键^{1,2} 陈浩^{1,3}

(¹ 南开大学社会学院社会心理学系, 天津 300350) (² 上海嗨优游测智能科技有限公司, 上海 200120) (³

广东省青少年心理健康与心态发展重点实验室, 广州 510275)

摘要 针对组织管理中“选拔表现—在岗绩效落差”的难题，本研究引入“阶段匹配”视角，采用基于游戏的测评，在真实组织情境下开展了两项独立实地研究，以检验特定认知能力的相对预测差异。研究1依托789名校招候选人样本，发现在时间紧迫、高压且信息新异的招聘选拔阶段(最佳表现情境)，工作记忆对录用结果的预测力相对贡献更高，凸显了快速信息加工的重要性。研究2以446名入职员工为样本，发现在长周期、多任务并行的在岗工作阶段(典型表现情境)，抑制控制对工作绩效的预测力相对贡献更高，反映了抗干扰和注意维持的关键作用。本研究呈现了特定认知能力在两个真实组织阶段中的相对预测差异，揭示了不同任务阶段对核心认知资源的差异化需求；细化了单一的一般能力预测观，为不同阶段的人才测评实践提供了实证参考。

关键词 招聘选拔, 工作绩效, 工作记忆, 抑制控制, 基于游戏的测评

分类号 B849: C93

1 引言

识才之要，在于具备洞察个体潜能的“慧眼”。如何在众多应聘者中精确甄别出未来的高绩效人才，始终是管理心理学与组织行为学关注的核心议题。为此，各类测评技术应运而生。其中，以认知能力测验为代表的标准化工具因其客观性和高效性，在企业组织规模化招聘选拔中被广泛采用(Schmidt & Hunter, 1998)。特别是一般心理能力(general mental

收稿日期: 2025年11月10日

* 中央高校基本科研业务费专项资金项目(63233147)资助

通讯作者: 陈浩, E-mail: hull@nankai.edu.cn

ability, GMA)测试, 凭借其对于培训、学习及工作相关结果的稳健预测力, 长期以来被确立为企业人员甄选领域的金标准(李英武 等, 2015; Schmidt & Hunter, 2004)。

然而, 在管理实践中, 企业组织常面临“选拔表现—在岗绩效落差”的困境: 部分在招聘选拔中脱颖而出的候选人, 入职后却未能成为高绩效员工(Pierce & Aguinis, 2013)。这一现象提示我们, 尽管 GMA 提供了基础的预测效度, 但仅依赖单一指标可能掩盖了特定认知能力(specific cognitive abilities)在不同任务中的差异化贡献(Krumm et al., 2014); 其绩效预测力可能随情境和任务阶段发生动态变化(Sackett et al., 2008)。基于此, 本研究旨在将特定认知能力作为 GMA 的过程性体现与功能分解, 深入探究其不同任务阶段的差异化预测效应。

1.1 基于心理测量与执行功能的认知能力预测效应

本文首先结合心理测量学与认知/执行功能的理论视角, 剖析所考察的认知能力。前者侧重刻画个体认知差异的潜在结构, 揭示认知能力在统计层面的组织方式; 后者则聚焦信息加工中具体认知资源的调用机制, 阐明其在动态过程中的运作逻辑(Carroll, 1993; Engle et al., 1999)。基于此, 本文借由“结构—过程”的互补性分析框架, 力图实现特定认知能力内在机理探析与预测效应检验的有机统一。

在心理测量学中, Spearman(1904)提出的 g 因子(general factor)构念, 反映了多项认知能力测验中的共同变异, 是 GMA 最主要的统计体现(Carroll, 1993)。凭借其跨任务的通用性, g 因子在预测学业成就、职业表现等方面展现出稳健的效度。然而, 正由于 g 因子侧重于捕捉能力测验间的共变成分, 常掩盖不同认知能力在预测特定效标时的独特贡献。为突破这一局限, 卡特尔-霍恩-卡罗尔模型(Cattell-Horn-Carroll model, CHC)提供了一个精细化的层级架构(McGrew, 2009)。该模型指出, 在顶层的 g 因子之下, 还存在流体智力(Gf)、短时记忆(Gsm)、视觉空间能力(Gv)等广义(认知)能力(broad abilities); 广义能力之下还有特定(认知)能力及指标(如工作记忆、快速旋转能力、空间扫描等), 它们在不同的任务情境中发挥着特异化的认知功能(Carroll, 1993; Schneider & Newman, 2015)。该模型结构引出一项关键议题: 针对不同情境和任务阶段的效标(如招聘录用、在岗绩效), 不同的特定认知能力的相对预测

权重是否存在系统性差异？厘清这一问题，不仅能在理论上划定不同认知能力的情境适用边界，亦能在实践中提升人员遴选的精确度与公平性。

在认知机制层面，执行功能及其子成分被视为个体在复杂情境中实现目标导向行为的认知控制过程(Diamond, 2016)。尽管在实证分析上整体执行功能与 g 因子高度关联，但二者在理论构念和分析范式上并不等同(Conway et al., 2003; Friedman & Robbins, 2022)。g 因子侧重于从测量学角度刻画跨任务的共同方差，反映认知能力的潜在特质结构；而执行功能则侧重于描述组织思维与调控行为的动态加工过程(Floyd et al., 2010)。这种交叠但可区分的逻辑同样适用于层级性的特定认知能力。例如，流体智力(Gf)在认知机制上高度依赖工作记忆以支持复杂推理(Draheim et al., 2022)，而知觉速度(Gs)则更多映射于信息加工效率。这意味着，特定认知能力的增量预测效应，反映了其作为 GMA 发挥作用时的具体加工路径，通过对认知资源的功能性分解、调用，精准契合具体情境任务的差异化需求。

职业情境中的实证研究为认知能力如何转化为实际绩效提供了证据：当测验范式所诱发的认知加工过程与目标岗位的关键工作特征具备功能同构性时，特定认知能力往往能提供独立于 g 因子的增量预测效度(Krumm et al., 2014)。例如，在飞行员选拔中，知觉速度测验对后续飞行训练表现的预测力优于通用智力量表，原因在于该测评工具精准捕捉了驾驶舱环境下所需的快速视觉扫描能力(Johnson et al., 2017)。同样，针对仓库工人的研究表明，知觉速度与准确性测验提供了超越 g 因子的增量解释力，这得益于该测验有效映射了重复性作业中所需的快速信息加工机制(Mount et al., 2008)。即便针对常规或低复杂度岗位，某些与 g 因子共享变异较低的特定能力指标(如视觉加工速度、心理反应时)，亦能通过表征个体底层的信息加工效率，在 g 因子之外进一步提升对工作绩效的预测效力(Lang et al., 2010; Nye et al., 2022)。

然而，仅基于静态的“任务—能力”匹配观尚不足以揭示预测效应的全貌。技能习得理论(Ackerman, 1988)表明，任务对认知资源的需求是动态变化的。具体到本文研究场景，招聘选拔往往表现为短时、高强度的信息加工情境，而在岗工作则更多体现为长周期、多干扰的常规作业环境。因此，有必要探明特定认知能力在选拔与在岗两个阶段的差异化预测效应。

1.2 阶段匹配视角下的特定认知能力差异化预测效应假设

基于上述任务—能力动态匹配的分析框架(即“阶段匹配”视角),本研究选取了四项兼具 g 因子高载荷且与职业绩效密切相关的特定认知能力,旨在探析和检验其在不同任务阶段的差异化预测效力。首先,工作记忆(working memory)作为执行功能的核心成分之一,负责对信息进行实时保持和操作整合(Conway et al., 2003; Diamond, 2016)。在 CHC 模型中,它被视为支撑 g 因子的关键认知资源,与短时记忆(Gsm)、流体智力(Gf)高度关联(Ackerman et al., 2005; Jewsbury et al., 2016; Kane et al., 2005)。研究表明,工作记忆在应对新异任务及早期学习时,往往能提供超越 g 因子的增量预测力(Alloway & Alloway, 2010),使其成为考察高认知负荷选拔情境下的关键能力指标。

其次,作为执行功能的另一关键成分,抑制控制(inhibitory control)主要负责冲突监控、目标维持和干扰屏蔽(Botvinick et al., 2001; Diamond, 2016; Eriksen & Eriksen, 1974)。在 CHC 模型中,它不仅与加工速度(Gs)相关(Jewsbury et al., 2016),也与短时记忆(Gsm)、视觉空间能力(Gv)共享部分变异(van Aken et al., 2019)。相较于工作记忆,抑制控制对真实世界中充满干扰的复杂实践领域具有广泛的预测价值(Draheim et al., 2022; St Clair-Thompson & Gathercole, 2006)。在长周期、多任务并行的在岗工作阶段,抑制控制可能发挥着比工作记忆更为核心的作用。

此外,为更全面地覆盖关键认知能力,本研究将空间表征(spatial ability)和路径规划(navigational planning)一并纳入考察范畴。空间表征的核心在于心理意象的生成与操纵(Shepard & Metzler, 1971),其主要映射于 CHC 模型的视觉加工(Gv)广义能力(Schneider & McGrew, 2018),并承载了部分流体智力(Gf)负荷(Varriale et al., 2018)。研究表明,以心理旋转为成分之一的空间表征能力可有效预测数字化情境下的人机交互表现(Dilling & Vogler, 2021; Pak et al., 2008)。路径规划则侧重于多步分析推理与策略生成(Sheppard & Cheatham, 2017)。它不仅反映执行功能中的计划与认知灵活性成分(Bocchi et al., 2017),也与流体智力(Gf)关联(Guglielmo et al., 2022)。这两项特定认知能力因其特异性的加工机制,预期将在不同任务阶段或情境中展现出差异化的预测价值,从而为解释个体表现贡献增量效度。

GMA 或 g 因子作为统摄性的上位能力, 虽然为个体在各类任务表现中提供了基础性的认知资源支撑(Schmidt & Hunter, 2004), 但其作用机制并非在所有情境下都是均质的。在从招聘选拔向在岗工作转换的过程中, 任务性质的变化导致了对特定认知加工资源的倚重有所差异。首先, 招聘选拔阶段构成了“最佳表现”情境(Sackett et al., 1988)。此时个体处于高唤醒状态, 需在短时间内调动认知资源以应对高强度的新异任务(Klehe & Anderson, 2007)。其核心特征是对受控注意和快速信息加工的高需求。面对陌生规则, 个体尚未形成自动化图式, 须调用 GMA 实时解码指令、构建表征并推理。尽管抑制控制、空间表征与路径规划等认知能力, 能协同支持推导与解析, 但在高压限时环境中, 整体表现的瓶颈在于工作记忆系统对信息进行暂时存储与加工的容量限制。作为支持流体智力(Gf)的底层机制, 工作记忆负责信息的暂存与加工(Kane et al., 2005)。在此情境下, 个体需不断更新认知集合以适应不断变化的规则, 这种对信息吞吐与整合的需求, 使得作为 GMA 主要过程性体现的工作记忆, 成为决定任务成败的关键区分性能力。因此, 本文预期在招聘选拔阶段, 工作记忆的预测力将相对高于其他特定认知能力。

其次, 在岗工作阶段则反映了“典型表现”情境(Sackett et al., 1988)。随着常规任务逐渐熟练化, 工作性质从短期集中的“高强度解析”转变为长周期、多目标的“持续性应对”。此阶段的挑战更多涉及复杂情境下的适应性调控, 对抑制控制的需求相对上升。虽然 GMA 及其他特定能力依然持续支撑信息处理, 但真实职场充满了动态干扰和目标冲突(如在处理本职工作时面临突发的跨部门协调、社交干扰或情绪压力等)。在此情境下, 绩效高低往往不取决于“反应得有多快”, 而取决于“做得有多稳”, 即个体能否长程维持目标导向并屏蔽干扰。抑制控制是个体主动屏蔽干扰并克服优势反应的能力(Diamond, 2013)。基于该冲突监控与资源调节机制(Botvinick et al., 2001), 高抑制控制的员工在面对干扰时, 可以更高效地维持目标并减少失误。因此, 本文预期在入职在岗阶段, 抑制控制的绩效预测相对贡献将高于其他特定认知能力。

综上所述, 尽管 GMA 构成了个体认知表现的共性基础, 但通过解析阶段性任务对认知资源的差异化需求与调用机制, 特定认知能力在“招聘选拔”和“入职在岗”两个真实组织情境中, 预期将表现出差异化的预测效力。据此, 本研究提出并检验以下三项假设: (1) 各

项认知能力测试在两个情境中均具有效标区分能力，即呈现显著组间差异：招聘选拔中，录用组优于未录用组；在岗工作中，高绩效组优于低绩效组。(2) 在招聘选拔阶段，对于录用结果的预测，工作记忆的相对贡献高于抑制控制、空间表征和路径规划。(3) 在入职在岗阶段，对于工作绩效的预测，抑制控制的相对贡献高于工作记忆、空间表征和路径规划。

1.3 基于游戏的测评与真实组织情境下的研究设计

为在同一测量框架下精准分析 GMA 及不同特定认知能力在不同任务阶段的相对重要性，并兼顾被试的卷入度与过程数据的获取，本研究采用基于游戏的测评(game-based assessment, GBA)作为核心工具(孙鑫 等, 2018; Landers et al., 2022)。相较于传统测验，GBA 将心理测量原理嵌入游戏化交互中，通常能提升个体参与沉浸感，并利用过程数据支持精细建模(Attali & Arieli-Attali, 2015; Gomez et al., 2022; Heinzen et al., 2015; Kiili et al., 2018; Köstering et al., 2015; Shute et al., 2016)。其数字化特性亦有助于提高测评效率与适应性(Leutner et al., 2023; Maruff et al., 2009; Pan et al., 2024; Pedersen et al., 2023; Sternberg et al., 2013)。本研究采用的 GBA 工具覆盖工作记忆、抑制控制、空间表征与路径规划四类典型的特定认知能力测试，在两项子研究中统一施测与计分。

尽管 GBA 具备精细化测量优势，但既有 GBA 研究多局限于低利害的实验室环境或非求职情境的学生样本，缺乏在真实组织生态中的系统性检验(Ramos-Villagrasa et al., 2022)。真实职场中，个体从“高压选拔”(最佳表现情境)向“长周期在岗”(典型表现情境)的转化是一个复杂的动态过程(Sackett et al., 1988)，现有实验室研究范式往往难以有效捕捉。这种现实情境维度的缺失，使得现有结论难以解释认知能力在不同阶段预测效度的变化模式，也限制了 GBA 研究的生态效度。

为此，本研究选取校园招聘及随后的在岗工作作为研究场景。相较于社会招聘，校招候选人年龄相近、先期工作经验较少且同质性较高，用人单位更倾向于发展式选拔(李彬, 白岩, 2020)，且招聘活动多在限定时段集中进行(许昉昉 等, 2019)。这些特征客观上提高了样本间的可比性，减少了由时间跨度与过往履历引起的测量误差和情境噪音。招聘选拔与在岗工作两个阶段在任务性质上的天然差异，构成了最佳表现和典型表现的对照情境，为考察认知能力在不同表现域下的差异化预测权重提供了理想的实验场。

针对真实选拔场景中未录用者绩效数据缺失的客观限制，本研究难以对同一初始样本进行全周期的纵向追踪，故采用分阶段的实地对照设计：研究 1 聚焦最佳表现情境，以校招候选人为样本，考察其 GBA 测验成绩与录用结果的关联；研究 2 则聚焦典型表现情境，以已正式入职的新员工为独立样本，通过回溯其在选拔阶段的 GBA 测验存档数据，探究认知能力与工作绩效的关系。该设计是针对组织情境中数据特征的权变策略：一方面，研究 1 基于初始求职者样本，保留了数据的未受限变异特征，能够有效检验认知能力对录用结果的预测效度；另一方面，研究 2 针对录用后的正式员工样本，通过回溯分析并以入职早期绩效为效标，既能确保与前期 GBA 测评的时间间隔可控，又能减少岗位调动和环境差异对长期绩效的噪音干扰(Murphy, 1989)。两项子研究采用统一的 GBA 测评工具，从而有效控制了测量方法变异的干扰。

2 研究 1：阶段匹配 I——工作记忆对录用结果的优先预测力

2.1 参与者

研究 1 的参与者为某世界 500 强企业校园招聘时的候选人，来自 170 所高校的 2024 届应届本硕毕业生。该企业初期收到逾千份简历，经资格审查后，有 803 人进入 GBA 测评环节完成作答。研究者对测试任务数据逐一审核，剔除不认真作答、各测试任务正确数低于平均值 3 个标准差的无效样本，最终获得有效样本 789 人。其中，男性 295 人(37.39%)、女性 494 人(62.61%)；平均年龄 23.36 岁($SD=1.25$)。参与者学科门类以经济学为主(554 人，70.22%)，同时涵盖管理学(90 人，11.41%)、工学(79 人，10.01%)等。测评环节之后，候选人继续进入企业的两轮综合面试评估环节，最终共有 101 人被录用，录用率为 12.80%。功效分析显示，在显著性水平 $\alpha=0.05$ 、统计功效 $1-\beta=0.80$ 且预期效应量为中等(Cohen's $d=0.50$)的条件下，录用组与未录用组至少各需 64 名参与者可满足检验要求。研究 1 的样本规模满足该标准。

2.2 流程

通过简历初筛的候选人在收到企业的测评邀请邮件后，登陆指定链接进入平台系统，参与线上 GBA 测评。该测评包含四项游戏化测验，参与者需依次完成。系统全程自动记录其操作行为与作答结果。为检验测评工具的信度，研究 1 设置了重测机制。既有研究表明，

在 1 天至数周的时间跨度内，个体间认知表现的差异不受时间间隔影响(Salthouse & Tucker-Drob, 2008)。据此，本研究要求参与者在首次测评结束后的 3 天内，自主选择时间再次登录平台系统进行重测。这一安排兼顾了应届生求职季的时间压力，赋予其必要的灵活性，避免了固定时间窗口带来的额外认知负担。在实际招聘选拔中，企业采用多阶段综合评估机制。简历初筛与 GBA 测评成绩仅作为初选阶段的参考指标，最终的录用决策主要依赖于后续的两轮面试(部门面试与高管面试)。面试环节由多名面试官综合评估候选人的问答表现、沟通表达及组织适配度，并据此做出最终录用决定。

2.3 材料

为提升招聘应用中的可操作性与生态效度，本研究基于经典心理学实验范式，开发了四项游戏化认知能力测试(如图 1 所示，详细说明见附录 1)，要求参与者按既定顺序依次完成。为便于后续分析与阐述，四项认知能力测试依次编号为：第一关(工作记忆测试)、第二关(空间表征测试)、第三关(抑制控制测试)和第四关(路径规划测试)。



图 1 基于游戏的认知能力测评界面示意图

第一关(工作记忆测试)：Corsi 积木测试任务(Corsi Block-Tapping Task)广泛应用于认知心理学和脑认知神经科学研究(Corsi, 1972; Milner, 1971)，是测量工作记忆的常用方法(Diamond, 2013)。该测试的完成主要依赖顶叶-前额叶皮层网络，涉及空间信息的存储与序

列加工(Kessels et al., 2000), 其数字化版本效度表现良好(Brunetti et al., 2014)。本研究在此范式中加入数值运算, 以检验个体对空间与数学信息整合加工的工作记忆容量。测试中, 参与者需识记短暂呈现的信封及其内部数字或算式, 并按数字大小与空间顺序将信封拖放至目标区。游戏共 10 轮且难度递增, 记录正确反应数及作答时长(毫秒)。

第二关(空间表征测试): 心理旋转测试任务(Mental Rotation Task)常用于评估个体的空间表征转换与操作能力, 广泛应用于空间认知研究(Shepard & Metzler, 1971; Vandenberg & Kuse, 1978), 其神经机制与顶叶皮层在处理空间关系时的激活相关(Zacks, 2008), 数字化版本信效度表现良好(Monahan et al., 2008)。本研究采用游戏化心理旋转测试评估空间表征能力, 要求参与者根据呈现的不规则立方体, 绘制其三视图以完成三维至二维的转换。游戏共 5 轮且难度递增, 记录正确反应数及作答时长(毫秒)。

第三关(抑制控制测试): Flanker 测试任务是评估抑制控制的经典工具(Diamond, 2013), 其神经机制涉及前扣带回与背外侧前额叶的协同作用, 要求参与者在存在干扰的情况下快速识别目标刺激, 抑制对干扰刺激的反应, 从而测量个体在认知冲突中的反应模式(Eriksen & Eriksen, 1974)。本研究采用游戏化 Flanker 测试评估个体抑制控制能力(Amin et al., 2023; Fan et al., 2002)。参与者需判断屏幕飞机方向并按键。中心与两侧一致时按中心方向, 不一致时按两侧方向; 两侧为云朵时仅按中心方向, 遇闪电时则抑制反应。游戏共 60 轮, 记录正确反应数及作答时长(毫秒)。

第四关(路径规划测试): 迷宫测试任务及其改编版常用于评估空间导航与规划能力。例如, Lind 等(2014)设计了虚拟岛屿测试, 以考察空间工作记忆和路径规划能力; Wang 等(2025)开发了三维空间搜索测试以评估空间认知。本研究采用游戏化迷宫测试评估路径规划能力。参与者需在地图上规划路径, 以最少步数到达终点。沿单一方向移动至障碍或边界停下计为一步; 停下后可转向或回退。游戏共 5 轮, 记录完成步数与最优步数之差, 以及作答时长(毫秒)。

在正式研究启动前, 研究者分批招募了多组受试者, 对上述测试的效标关联效度进行了检验。结果显示效度良好, 具体详见附录 2。

2.4 数据分析

研究 1 使用 Python 3.8 进行数据处理与分析。各测试的计算指标基于核心表现设定：工作记忆测试、空间表征测试、抑制控制测试皆以正确(反应)数、作答时长为表现指标；路径规划测试以最优步数差、作答时长为表现指标。各指标中，正确数为正向指标，即正确数越高，作答表现越好；作答时长、最优步数差为负向指标，即作答时长越低、最优步数差越低，作答表现越好。研究者首先计算同一测试中各指标的多轮次均值，作为各指标测评得分；随后对同一参与者初测和重测的指标得分进行重测信度检验。在此基础上，研究者再对两次测评的各指标得分分别求取均值，并做标准化处理，作为参与者在四项认知能力测试上的最终表现，用于后续分析。接着，研究者采用独立样本 t 检验比较录用与未录用组之间的表现差异，并使用逻辑回归和 XGBoost 算法比较各认知能力测评指标对录用结果的预测力。

2.5 结果

2.5.1 结构效度检验

为检验结构效度，研究者首先对研究 1 的数据进行因子分析适用性检验。Bartlett 球形检验($\chi^2 = 988.58, p < 0.001$)和 KMO 检验(0.76)结果表明数据适合进行因子分析。随后，研究者构建了验证性因子分析模型。初步分析与既有游戏化测评结构效度研究发现类似(Landers et al., 2022)：由于测试关卡需要维持必要的游戏性，指标间难以避免地存在载荷交叉，比如第二关心理旋转、第四关迷宫都涉及对呈现的视觉信息进行心理层面的加工处理，这种关卡特征和游戏化的开发导向，使得基于经典测验的 CFA 模型难以获得理想的拟合($\chi^2(20) = 152.57, p < 0.001, CFI = 0.86, RMSEA = 0.09$)。参照 Landers et al. (2022)关于游戏化测评结构效度检验的处理建议，研究者在允许部分测试指标残差相关的前提下对模型进行重新拟合，模型拟合度得到改善($\chi^2(17) = 68.74, p < 0.001, CFI = 0.95, RMSEA = 0.06$)。

为进一步检验测评是否能区分一般认知能力与特定认知能力，研究 1 采用 Schmid-Leiman 方法的正交层级模型分析测评结构(Schmid & Leiman, 1957)。该方法是处理层级因子结构的经典方法(Jensen, 1998)，可将测试的方差分解为互相独立的部分。结果显示，在基于四类认知能力测试构建的正交层级模型中， g 因子作为第一因子，解释共同方差占比为 48.02%。与此同时，各测评指标在其对应的下位因子上仍保留载荷，支持该测评不仅能

够反映一般认知能力，也包含特定认知能力的成分，为后续检验它们的差异化预测作用提供了结构基础。以上结果表明，研究 1 所使用的 GBA 结构效度总体可接受。

2.5.2 重测信度检验

研究 1 采用组内相关系数(intraclass correlation coefficient, ICC)计算指标的变异性与总变异性之比。鉴于测评次段(首测、重测)为固定效应，参与者为随机效应，且主要关注差异的稳定性，本研究采用双向混合效应模型的一致性类型进行计算(即 ICC(3,1)模型)，结果见表 1。参考 Koo 和 Li(2016)的标准：ICC > 0.90 为优秀，0.75 ~ 0.90 为良好，0.50 ~ 0.75 为中等，< 0.50 为不足。结果显示，各测试任务在两次测评之间的表现总体一致(ICC=0.66~0.85)，表明这些认知能力测试在重复测量上具有可接受的一致性。

表 1 研究 1 重测信度检验

测评指标特征	首测	重测	ICC (3,1) 95% CI
工作记忆测试正确数	31.33 (6.51)	31.32 (6.39)	0.85 [0.83, 0.87]
空间表征测试正确数	9.24 (2.90)	9.22 (2.86)	0.77 [0.74, 0.80]
抑制控制测试正确数	40.11 (10.00)	40.04 (10.14)	0.70 [0.66, 0.73]
路径规划测试最优步数差	4.18 (1.60)	4.29 (1.45)	0.69 [0.65, 0.72]
工作记忆测试作答时长	79572.18 (11482.44)	81922.86 (10238.81)	0.72 [0.65, 0.77]
空间表征测试作答时长	229169.29 (29949.10)	229215.36 (29828.25)	0.72 [0.68, 0.75]
抑制控制测试作答时长	70149.56 (4911.96)	70145.75 (5102.04)	0.84 [0.82, 0.86]
路径规划测试作答时长	133966.39 (59777.30)	133862.40 (59465.98)	0.66 [0.62, 0.70]

注：正确数越高、最优步数差越低、作答时长越低(即反应速度越快)，作答表现越好，下同。

2.5.3 独立样本 *t* 检验

研究 1 采用独立样本 *t* 检验比较录用组与未录用组在四项测试上的表现，结果见表 2。录用组的工作记忆测试正确数(0.96 ± 0.85)优于未录用组(-0.14 ± 0.94)，差异显著($p < 0.001$)。在空间表征测试正确数上，录用组(0.69 ± 0.77)同样优于未录用组(-0.10 ± 0.99)，差异显著($p < 0.001$)。录用组的抑制控制测试正确数(0.74 ± 0.67 vs. 0.11 ± 1.00)、路径规划测试最优步数差(-0.62 ± 0.84 vs. 0.09 ± 0.99)，也显著更佳($ps < 0.001$)。

表 2 研究 1 独立样本 t 检验

特征	未录用组	录用组	t	p	Cohen's d	95% CI
	($M \pm SD$)	($M \pm SD$)				
工作记忆测试正确数	-0.14 (0.94)	0.96 (0.85)	12.03	< 0.001	1.23	[1.01, 1.45]
空间表征测试正确数	-0.10 (0.99)	0.69 (0.77)	9.25	< 0.001	0.89	[0.68, 1.11]
抑制控制测试正确数	-0.11 (1.00)	0.74 (0.67)	11.02	< 0.001	1.00	[0.78, 1.22]
路径规划测试最优步数差	0.09 (0.99)	-0.62 (0.84)	-7.71	< 0.001	-0.77	[-0.99, -0.56]
工作记忆测试作答时长	-0.02 (1.02)	0.10 (0.85)	1.28	0.203	0.13	[-0.09, 0.34]
空间表征测试作答时长	0.03 (1.00)	-0.18 (0.98)	-1.96	0.052	-0.21	[-0.42, 0.01]
抑制控制测试作答时长	-0.00 (0.98)	0.01 (1.12)	0.05	0.959	0.01	[-0.21, 0.22]
路径规划测试作答时长	-0.04 (0.98)	0.29 (1.10)	2.86	0.005	0.32	[0.11, 0.53]

2.5.4 逻辑回归分析

鉴于样本中分组数量不平衡，研究 1 采用类别权重的自动调整，以减轻不平衡对模型拟合的影响，使用 L2 正则化用于评估模型稳健性，分测验模型(ROC AUC = 0.84)表现优于 g 因子模型(ROC AUC = 0.69)。在控制参与者年龄、性别与学科门类后，将不同测试表现指标纳入逻辑回归模型，结果见表 3。依据 Chen 等人(2010)的标准，g 因子模型中，g 因子($B = 0.53$, $SE = 0.09$, $OR = 1.69$, $p < 0.001$)呈显著小效应。分测验模型中，工作记忆测试正确数预测力相对更强($B = 0.97$, $SE = 0.16$, $OR = 2.64$, $p < 0.001$)，接近中等效应；抑制控制测试正确数($B = 0.64$, $SE = 0.19$, $OR = 1.90$, $p < 0.001$)与空间表征测试正确数($B = 0.43$, $SE = 0.18$, $OR = 1.53$, $p = 0.020$)呈显著小效应，路径规划测试最优步数差($B = -0.51$, $SE = 0.17$, $OR = 0.60$, $p = 0.002$)亦呈显著小效应。

表 3 研究 1 逻辑回归模型

模型与特征 ¹	<i>B</i>	<i>SE</i>	Wald	<i>p</i>	OR	95% CI
模型 1						
截距	-0.43	2.05	0.04	0.835	0.65	[0.01, 36.54]
年龄	-0.07	0.09	0.58	0.447	0.94	[0.79, 1.11]
性别	0.05	0.23	0.04	0.840	1.05	[0.67, 1.65]
学科-经济学	-0.23	0.23	0.92	0.337	0.80	[0.51, 1.26]
g 因子	0.53	0.09	37.88	< 0.001	1.69	[1.43, 2.00]
模型 2						
截距	0.74	2.32	0.10	0.749	2.10	[0.02, 198.35]
年龄	-0.15	0.10	2.30	0.129	0.86	[0.71, 1.05]
性别	0.13	0.26	0.23	0.629	1.13	[0.68, 1.89]
学科-经济学	-0.29	0.27	1.21	0.271	0.75	[0.44, 1.26]
工作记忆测试正确数	0.97	0.16	35.23	< 0.001	2.64	[1.92, 3.65]
工作记忆测试作答时长	0.24	0.14	2.87	0.090	1.28	[0.96, 1.69]
空间表征测试正确数	0.43	0.18	5.39	0.020	1.53	[1.07, 2.19]
空间表征测试作答时长	0.16	0.15	1.26	0.261	1.18	[0.89, 1.57]
抑制控制测试正确数	0.64	0.19	11.21	< 0.001	1.90	[1.31, 2.77]
抑制控制测试作答时长	-0.02	0.12	0.03	0.856	0.98	[0.78, 1.23]
路径规划测试最优步数差	-0.51	0.17	9.59	0.002	0.60	[0.43, 0.83]
路径规划测试作答时长	0.10	0.12	0.72	0.395	1.11	[0.88, 1.40]

2.5.5 XGBoost 模型验证

为检验不同测试表现指标构建模型的稳健性，研究 1 采用 XGBoost 分类模型并结合 SHAP 方法分析各表现指标对效标预测的贡献。XGBoost 是一种基于梯度提升树的机器学习算法，能捕捉变量间非线性复杂关系(Chen & Guestrin, 2016)，以弥补传统回归模型在模式拟合上的潜在不足。模型结果输出 SHAP 值，用于衡量单个个案中各指标特征的贡献大小与方向；其中 SHAP 值的绝对值反映影响强度，符号正负分别表示对预测结果的正向或负向作用。

研究 1 采用 5 折交叉验证，并使用 SMOTE-Tomek 对类别进行重采样以实现平衡；整体性能良好(Accuracy = 0.82, ROC AUC = 0.83)。如图 2 所示，基于 SHAP 值衡量的特征

¹ 注：为验证特定能力的增量贡献并进行稳健性检验，研究 1 额外构建了“g + 4 个残差化特定因子”模型。特定因子采用效率分 Rate Correct Score(正确数除以作答时长，或最优步数差除以作答时长)指标，并通过回归分析剔除 g 因子变异(取残差项)。结果表明，特定因子整体为模型提供了显著的增量解释力($\Delta\chi^2 = 13.03$, $p = 0.011$)。

贡献大小与逻辑回归分析结果基本一致，工作记忆测试正确数的 SHAP 值为 0.67，高于抑制控制测试正确数(0.42)、路径规划测试最优步数差(0.36)及空间表征测试正确数(0.25)。这表明，在测试表现指标中，工作记忆测试正确数对录用结果预测的影响相对更强。

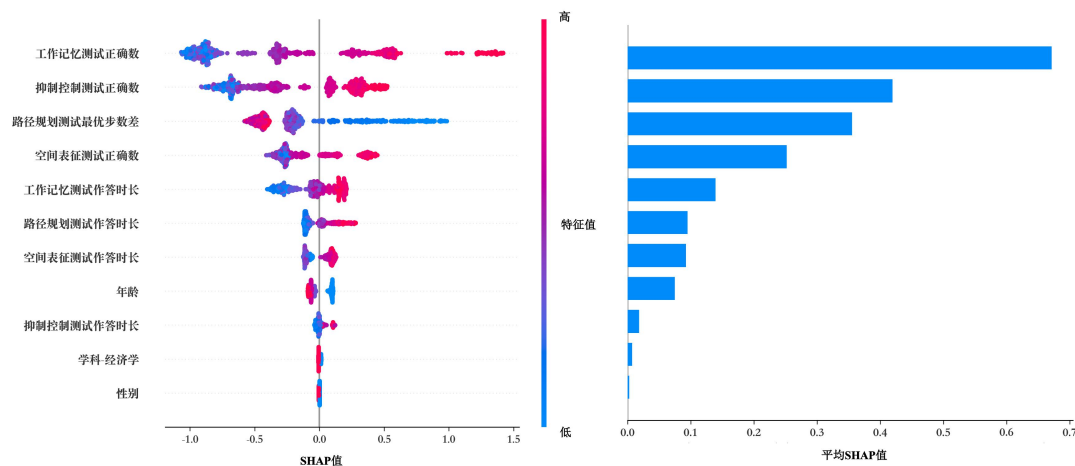


图2 研究1各特征 SHAP 值

2.6 讨论

基于“阶段匹配”视角，研究1的分析结果支持假设1与假设2。在认知能力测评表现上，录用组总体优于未录用组(假设1成立)。为细致比较各能力测试指标的预测力，研究1采用逻辑回归模型与XGBoost算法进行了深入分析。两类模型分析得到趋同结果：工作记忆相对贡献更高(假设2成立)。该发现与保持、更新和整合等执行功能过程在高利害且时间紧迫的选拔任务阶段更为关键的理论预期相一致。通过对比g因子模型与分测验模型，研究1还发现尽管g因子作为一般心理能力的指标具有显著的预测作用，但纳入特定认知能力的分测验模型展现出了更优的预测效能与增量解释力。

研究1检验了认知能力测评反映的GMA在录用结果上的普遍预测力，发现工作记忆在招聘录用阶段的预测作用相对贡献更高。研究2将在“阶段匹配”视角下，转向对入职后绩效的预测分析。

3 研究2：阶段匹配II——抑制控制对在岗绩效的优先预测力

3.1 参与者

研究 2 于 2023 至 2024 年间在 4 家世界 500 强及行业龙头企业的正式入职员工(曾经历 GBA 招聘测试)工作情境中开展, 最终纳入分析的有效参与者为 446 人; 其中, 男性 177 人(39.69%)、女性 269 人(60.31%), 平均年龄 23.92 岁($SD = 1.46$)。参与者分别来自 4 家企业: A 企业 34 人(7.62%)、B 企业 119 人(26.68%)、C 企业 175 人(39.24%)和 D 企业 118 人(26.46%)。参与者的主要岗位类型为管理培训类(154 人, 34.53%), 其余涵盖营销类(82 人, 18.39%)、技术类(70 人, 15.70%)及其他岗位。依据来自企业的绩效评定结果, 其中 280 人被归为高绩效组(62.78%), 166 人被归为低绩效组(37.22%)。功效分析显示, 在显著性水平 $\alpha = 0.05$ 、统计功效 $1 - \beta = 0.80$ 且预期效应量为中等(Cohen's $d = 0.50$)的条件下, 高绩效组与低绩效组至少各需 64 名参与者可满足检验要求。研究 2 的样本规模满足该标准。

3.2 流程与分析

研究 2 所有参与者都曾在校园招聘阶段完成 GBA 测试, 后被正式录用, 分别从事管理培训、营销、技术等不同岗位工作, 并在入职 3 至 6 个月期间参与所在企业的正式绩效考核。由于员工参与者所在岗位类型多样, 且不同企业的绩效考核体系存在差异, 绩效量纲无法统一。因此, 研究 2 在收集员工绩效信息时, 要求企业依据内部绩效考核标准对候选人表现进行分类, 将其划分为“高绩效”与“低绩效”两类。研究 2 沿用了研究 1 的分析方法。

3.3 结果

3.3.1 结构效度检验

研究 2 再次针对 GBA 测评工具进行结构效度检验, 方法与研究 1 相同。在允许部分测试指标残差相关的前提下进行拟合后, 拟合指标得到改善, 结构效度总体可接受; 详细内容见附录 3。

3.3.2 独立样本 t 检验

研究 2 采用独立样本 t 检验比较高绩效组与低绩效组在四项能力测试上的表现, 结果见表 4。高绩效组的工作记忆测试正确数(0.19 ± 0.90)、作答时长(-0.14 ± 0.99)高于低绩效组(-0.33 ± 1.08 ; 0.24 ± 0.97), 差异显著($ps < 0.001$)。在空间表征测试正确数、作答时长上, 高绩效组(0.20 ± 0.87 ; -0.13 ± 1.01)同样优于低绩效组(-0.33 ± 1.12 ; 0.22 ± 0.95), 差异显著($ps <$

0.001)。高绩效组的抑制控制测试正确数(0.26 ± 0.83 vs. -0.44 ± 1.11)、路径规划测试最优步数差(-0.19 ± 0.91 vs. 0.33 ± 1.07)，也表现显著更佳($ps < 0.001$)。

表 4 研究 2 独立样本 *t* 检验

特征	低绩效 ($M \pm SD$)	高绩效 ($M \pm SD$)	<i>t</i>	<i>p</i>	Cohen's <i>d</i>	95% CI
工作记忆测试正确数	-0.33 (1.08)	0.19 (0.90)	5.211	< 0.001	0.52	[0.33, 0.72]
空间表征测试正确数	-0.33 (1.12)	0.20 (0.87)	5.252	< 0.001	0.53	[0.33, 0.73]
抑制控制测试正确数	-0.44 (1.11)	0.26 (0.83)	7.087	< 0.001	0.72	[0.52, 0.92]
路径规划测试最优步数差	0.33 (1.07)	-0.19 (0.91)	-5.221	< 0.001	-0.52	[-0.72, -0.33]
工作记忆测试作答时长	0.24 (0.97)	-0.14 (0.99)	-3.932	< 0.001	-0.38	[-0.58, -0.19]
空间表征测试作答时长	0.22 (0.95)	-0.13 (1.01)	-3.602	< 0.001	-0.35	[-0.55, -0.16]
抑制控制测试作答时长	0.02 (1.00)	-0.01 (1.01)	-0.363	0.717	-0.04	[-0.23, 0.16]
路径规划测试作答时长	0.10 (1.01)	-0.06 (0.99)	-1.595	0.112	-0.16	[-0.35, 0.04]

3.3.3 逻辑回归分析

沿用研究 1 分析策略，研究 2 在预测效度评估中引入 L2 正则化以确保模型稳健性，结果发现分测验模型(ROC AUC = 0.76)表现优于 g 因子模型(ROC AUC = 0.71)。在控制参与者年龄、性别、岗位后，将不同测试的表现指标纳入逻辑回归模型，结果见表 5。依据 Chen 等人(2010)的标准，g 因子模型中，g 因子($B = 0.47, SE = 0.07, OR = 1.60, p < 0.001$)呈显著小效应。分测验模型中，抑制控制测试正确数的预测作用相对更强($B = 0.58, SE = 0.13, OR = 1.78, p < 0.001$)，呈显著小效应；空间表征测试正确数($B = 0.37, SE = 0.12, OR = 1.45, p = 0.002$)、路径规划测试最优步数差($B = -0.27, SE = 0.12, OR = 0.76, p = 0.023$)，与工作记忆测试作答时长($B = -0.26, SE = 0.12, OR = 0.77, p = 0.032$)，均呈显著小效应。

表 5 研究 2 逻辑回归模型

模型与特征 ²	<i>B</i>	<i>SE</i>	Wald	<i>p</i>	OR	95% CI
模型 1						
截距	1.71	1.73	0.97	0.324	5.51	[0.19, 164.21]
年龄	-0.04	0.07	0.32	0.571	0.96	[0.83, 1.11]
性别	-0.21	0.22	0.90	0.342	0.81	[0.53, 1.24]
岗位-管培类	-0.20	0.22	0.77	0.381	0.82	[0.53, 1.27]
g 因子	0.47	0.07	46.72	< 0.001	1.60	[1.40, 1.83]
模型 2						
截距	0.04	1.83	0.00	0.985	1.04	[0.03, 37.49]
年龄	0.03	0.08	0.11	0.743	1.03	[0.88, 1.19]
性别	-0.15	0.23	0.41	0.521	0.86	[0.55, 1.35]
岗位-管培类	0.09	0.24	0.13	0.723	1.09	[0.68, 1.75]
工作记忆测试正确数	0.24	0.12	3.68	0.055	1.27	[1.00, 1.61]
工作记忆测试作答时长	-0.26	0.12	4.61	0.032	0.77	[0.61, 0.98]
空间表征测试正确数	0.37	0.12	9.66	0.002	1.45	[1.15, 1.84]
空间表征测试作答时长	-0.02	0.12	0.03	0.857	0.98	[0.77, 1.24]
抑制控制测试正确数	0.58	0.13	20.28	< 0.001	1.78	[1.39, 2.30]
抑制控制测试作答时长	-0.15	0.11	1.74	0.187	0.86	[0.69, 1.08]
路径规划测试最优步数差	-0.27	0.12	5.16	0.023	0.76	[0.60, 0.96]
路径规划测试作答时长	-0.14	0.12	1.44	0.230	0.87	[0.69, 1.09]

3.3.4 XGBoost 模型验证

为检验不同测试表现指标构建模型的稳健性，研究 2 继续采用 XGBoost 分类模型并结合 SHAP 方法分析各表现指标对效标预测的贡献。研究具体采用 5 折交叉验证，并使用超参数搜索确定参数选择，整体性能良好(Accuracy = 0.71, ROC AUC = 0.74)。如图 3 所示，基于 SHAP 值衡量的特征贡献大小与逻辑回归分析结果基本一致，抑制控制测试正确数的 SHAP 值为 0.32，高于路径规划测试最优步数差(0.25)、工作记忆测试作答时长(0.16)以及空间表征测试正确数(0.15)。以上结果表明，在所有测试表现指标中，抑制控制测试正确数对绩效预测结果的影响相对更强。

² 注：为验证特定能力的增量贡献并进行稳健性检验，研究 2 中同样额外构建了“g + 4 个残差化特定因子”模型。特定因子采用效率分 Rate Correct Score 指标(构建同研究 1)，并通过回归分析剔除 g 因子变异(取残差项)。结果表明，特定因子整体为模型提供了显著的增量解释力($\Delta\chi^2 = 10.34, p = 0.035$)。

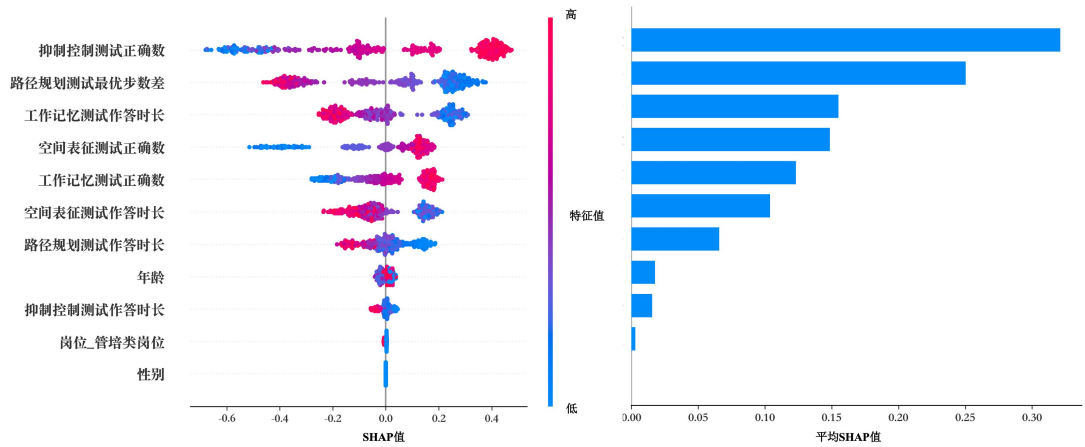


图3 研究2各特征 SHAP 值

3.4 讨论

研究2基于“阶段匹配”视角，进一步考察了个体认知能力与进入职场后绩效表现之间的关系。独立样本 t 检验分析表明，高绩效组在工作记忆、空间表征、抑制控制、路径规划这四类认知能力测试的表现显著优于低绩效组，验证了 GMA 是预测工作表现的重要因素 (Schmidt & Hunter, 2004)，支持假设 1 的成立。

逻辑回归分析显示，不同于研究 1，工作记忆测试相应指标在进入常规工作阶段后的预测力排位相对下降，抑制控制测试正确数在预测绩效时的作用则相对贡献更高。为了进一步检验该结果，研究 2 采用 XGBoost 分类模型结合 SHAP 方法分析各测试指标对绩效预测的相对贡献。结果验证了该发现的稳健性，支持假设 3 的成立。Flanker 测试被广泛认为是评估抑制控制与冲突监控的经典范式，其核心机制在于个体在面对干扰时能够抑制无关信息并保持对目标的专注。在实际工作环境中，这种能力关乎员工能否高效、稳定地完成各项工作任务，并避免因干扰而出现错误。这一发现呼应并拓展了 Murphy(1989)的阶段绩效理论，个体在不同任务阶段对不同的认知能力和执行功能过程的依赖存在差异。

最后，模型比较结果显示，g 因子具有一定预测效力，但分测验模型优于仅包含 g 因子的模型。这表明，在复杂的在岗绩效预测中，除了一般认知能力，特定认知能力能提供增量预测力。

4 总讨论

本研究采用基于游戏的测评和高生态效度设计，开展了两项独立实地对照研究，旨在揭示特定认知能力在两个真实组织阶段中的相对预测差异。本研究首先验证了测评工具的区别效度：录用组的认知表现显著优于未录用组，高绩效组亦显著优于低绩效组(支持假设 1)。更为重要的是，研究进一步呈现了两个任务阶段中特定认知能力的相对预测差异：招聘选拔阶段，工作记忆对录用结果的相对贡献更高(支持假设 2)；入职在岗阶段，抑制控制则对工作绩效展现出相对更强的解释力(支持假设 3)。上述发现为深入理解认知能力与工作表现的关系提供了新的实证证据，并带来重要的理论与实践启示。

4.1 理论意义

本研究结合心理测量学与认知/执行功能理论，以“结构—过程”互补性分析框架为基础，引入“阶段匹配”视角，剖析和实证检验了认知资源在不同任务阶段的差异化调用机制。这一工作不仅在一定程度上实现了特定认知能力内在机理探析与预测效应考察的有机统一，拓展了能力—绩效兼容性原则(Schneider & Newman, 2015)，也深化了 Murphy(1989)的阶段性能效理论。以往研究多视 GMA 为预测工作绩效的通用稳定指标(Schmidt & Hunter, 2004)，而相对忽视了不同阶段任务对认知资源的特异性需求与调用。本研究发现，在 CHC 模型视野下，虽然特定认知能力共享 GMA 基础，但其在选拔和在岗阶段呈现出功能分化。这表明，将研究视角从宏观的 GMA 细化至具体的认知/执行功能成分，能更精准地揭示出认知能力对不同阶段表现的内在作用机制。

其次，本研究解析了招聘选拔作为最佳表现情境(Sackett et al., 1988)，对工作记忆及其相关流体智力成分的优先调用。在此阶段，个体需面对高新异性、时间紧迫且高认知负荷的评估任务。基于 CHC 模型和执行功能理论，工作记忆紧密关联流体智力(Gf)与短时记忆(Gsm)，是制约个体信息加工效率的核心限制因素(Ackerman, 1988; Schneider & McGrew, 2018)。本研究发现工作记忆在录用预测中的作用相对更高，这与 Broadbent 等(2023)的观点相呼应，工作记忆容量较高的个体能够通过高效的执行注意和信息更新，做出更准确的判断。

此外，本研究揭示了在岗工作作为典型表现情境(Sackett et al., 1988)，对抑制控制及其相关行为调控机制的依重。与选拔阶段不同，入职后绩效不再仅依赖于快速学习，更取决

于个体在长时间、多线程、高干扰环境下的持续自主调节能力。Hofmann 等(2012)指出, 执行功能的三大核心成分共同构成了个体自我调节的认知基础。本研究进一步发现, 抑制控制对在岗绩效的解释力相对更高。这一结果提示, 个体屏蔽无关信息、维持专注的能力, 可能为长期的自我调节与绩效稳定提供了关键的认知基础。此外, 这一发现也为理解自我控制(self-control)——这一常被归为非认知领域特质——与绩效间关系的认知基础提供了新证据和新视角(Troll et al., 2022)。以往研究表明, 自我控制能显著预测职业成功等重要结果(Duckworth & Gross, 2014; Moffitt et al., 2011)。本研究进一步从认知机制层面表明, 这种预测力可能源于抑制控制作为核心执行功能, 支撑了个体在复杂工作场景下的行为稳定性与抗干扰性。

4.2 方法论贡献与实践启示

首先, 本研究检视了 GBA 在真实组织情境下的应用价值与跨文化适用性。该工具基于经典心理学范式设计, 重测信度优异(Enkavi et al., 2019), 预测效应良好, 在心理测量新工具研发与现实工作场景应用之间架起了桥梁(Allen et al., 2024)。本研究结果呼应了 Landers 等(2022)关于 GBA 在公平性和效度方面具有潜在优势的观点, 并弥补了现有研究多依赖在校样本、缺乏真实在岗绩效证据的局限(Woods et al., 2020)。此外, 本研究通过逻辑回归与机器学习方法的双重验证, 证实了该工具在中国企业情境中的预测稳健性, 为 GBA 在非西方文化背景下的应用提供了实证支持, 拓展了其跨文化适用边界(徐俊怡, 李中权, 2021)。

其次, 本研究在两个真实组织阶段中所发现的差异化结果, 可为企业在不同阶段的能力测评实践提供参考。尽管认知能力测评无法解释工作绩效的全部变异, 但其能够提供一一个相对稳定且客观的基线, 有效规避传统面试等选拔方式中潜藏的系统性偏差(如光环效应、刻板印象等; Judge & Cable, 2004; Thorndike, 1920), 从而提升人才配置的效率与公平性。

4.3 研究局限与展望

本研究在两个真实组织情境下考察了特定认知能力的相对预测差异, 受制于实地研究的客观条件, 存在以下局限, 亦为未来研究提供了深化方向。

首先, 受限于未录用者绩效数据缺失的客观现实, 本研究采用了两个独立样本的阶段实地对照设计。这一设计能够有效呈现不同组织阶段的相对预测差异, 但无法在同一群体

内部直接比较两阶段的认知预测效应，个体层面的阶段内变异有待未来纵向研究进一步检验。此外，两项子研究样本分别为处于职业早期的应届生和新员工，其较高的职业可塑性可能限制了结论的外部效度。未来研究可在数据和现实条件允许的情况下，采用贯通式纵向设计，并将样本拓展至不同职业发展阶段的群体，以考察经验积累对认知能力预测效度的调节作用。

其次，出于实地测评效率与可操作性的考量，本研究仅聚焦工作记忆、抑制控制等核心执行功能。随着游戏化测评技术的成熟，未来可开发更为精细的测验工具，构建完整的认知能力图谱，以全面检验不同细分能力对复杂工作绩效的增量贡献。

再次，对比招聘选拔和入职在岗两个独立阶段研究的发现，认知能力的整体解释力有所下降。这间接表明随着任务熟练度提升，员工对认知资源的依赖可能逐渐减弱(Murphy, 1989)；同时，在岗绩效的多维性(Tett & Burnett, 2003)也使得非认知因素的作用日益凸显。未来研究应在认知能力之外进一步纳入人格特质(Duckworth et al., 2019; Roberts et al., 2007; Wilmot & Ones, 2019)、动机、情绪调节(Beilock & Carr, 2005; De Dreu et al., 2008)及工作自主性(Morgeson et al., 2005)等因素，构建更为立体的职业表现预测框架。

最后，本研究发现主要基于中国企业组织情境，而文化差异可能塑造不同的认知加工模式(如整体性思维和分析性思维; Nisbett et al., 2001)。未来需在更广泛的文化背景下检验本研究结论的稳健性，探讨文化因素在认知能力与工作绩效间的调节作用，从而丰富并发展能力测评的跨文化适用性理论。

5 结论

(1) 在招聘选拔与入职在岗两个阶段，认知能力测评的整体表现方面，录用组优于未录用组，高绩效组优于低绩效组。

(2) 在招聘录用预测中，相较于抑制控制、空间表征和路径规划，工作记忆的相对贡献更高。

(3) 在入职后绩效预测中，相较于工作记忆、空间表征和路径规划，抑制控制的相对贡献更高。

参考文献

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs?. *Psychological Bulletin*, *131*(1), 30–60.
- Allen, K., Brändle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., ... & Schulz, E. (2024). Using games to understand the mind. *Nature Human Behaviour*, *8*(6), 1035–1043.
- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29.
- Amin, R., Tat, P., & Seitz, A. (2023). Validation of a Virtual Reality Flanker Task. *UC Riverside Undergraduate Research Journal*, *17*(1), 11–19.
- Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education*, *83*, 57–63.
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and “choking under pressure” in math. *Psychological Science*, *16*(2), 101–105.
- Bocchi, A., Carrieri, M., Lancia, S., Quaresima, V., & Piccardi, L. (2017). The Key of the Maze: The role of mental imagery and cognitive flexibility in navigational planning. *Neuroscience Letters*, *651*, 146–150.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.
- Broadbent, D. P., D'Innocenzo, G., Ellmers, T. J., Parsler, J., Szameitat, A. J., & Bishop, D. T. (2023). Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *92*, 121–132.
- Brunetti, R., Del Gatto, C., & Delogu, F. (2014). eCorsi: Implementation and testing of the Corsi block-tapping task for digital tablets. *Frontiers in Psychology*, *5*, Article 939.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA* (pp. 785–794). ACM.
- Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation*, *39*(4), 860–864.
- Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.
- Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain*. McGill University.
- De Dreu, C. K., Nijstad, B. A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review*, *12*(1), 22–49.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.
- Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 11–43). American Psychological Association.
- Dilling, F., & Vogler, A. (2021). Fostering spatial ability through computer-aided design: A case study. *Digital Experiences in Mathematics Education*, *7*, 323–336.
- Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*, *29*(4), 1143–1197.
- Duckworth, A. L., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science*, *23*(5), 319–325.
- Duckworth, A. L., Quirk, A., Gallop, R., Hoyle, R. H., Kelly, D. R., & Matthews, M. D. (2019). Cognitive and noncognitive predictors of success. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *116*(47), 23499–23504.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

- Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5472–5477.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.
- Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan executive function system and the Woodcock–Johnson III tests of cognitive abilities. *Psychology in the Schools*, *47*(7), 721–738.
- Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), 72–89.
- Gomez, M. J., Ruipérez-Valiente, J. A., & Clemente, F. J. G. (2022). A systematic literature review of game-based assessment studies: Trends and challenges. *IEEE Transactions on Learning Technologies*, *16*(4), 500–515.
- Guglielmo, G., Veld, E. H. I. T., Klineciewicz, M., & Spronk, P. (2022). Out of the maze: Investigating Fluid Intelligence and Numeracy as Predictive Factors of Planning Skills Using Video Games. In *11th International Conference on Games and Learning Alliance (GALA 2022), Tampere, Finland* (pp. 202–211). Springer International Publishing.
- 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. Wood (Eds.), *Gamification in Education and Business* (pp. 201–207). Springer.
- Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*(3), 174–180.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

- Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell—Horn—Carroll model. *Journal of Experimental Psychology: General, 145*(2), 220–245.
- Johnson, J. F., Barron, L. G., Carretta, T. R., & Rose, M. R. (2017). Predictive validity of spatial ability and perceptual speed tests for aviator training. *The International Journal of Aerospace Psychology, 27*(3-4), 109–120.
- Judge, T. A., & Cable, D. M. (2004). The effect of physical height on workplace success and income: Preliminary test of a theoretical model. *Journal of Applied Psychology, 89*(3), 428–441.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71.
- Kessels, R. P., Van Zandvoort, M. J., Postma, A., Kappelle, L. J., & De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Applied Neuropsychology, 7*(4), 252–258.
- Kiili, K., Moeller, K., & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training—In-game metrics as learning indicators. *Computers & Education, 120*, 13–28.
- Klehe, U.-C., & Anderson, N. (2007). Working hard and working smart: Motivation and ability during typical and maximum performance. *Journal of Applied Psychology, 92*(4), 978–992.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163.
- Köstering, L., Schmidt, C. S. M., Egger, K., Amtage, F., Peter, J., Klöppel, S., Beume, L.-A., Hoeren, M., Weiller, C., & Kaller, C. P. (2015). Assessment of planning performance in clinical samples: Reliability and validity of the Tower of London task (TOL-F). *Neuropsychologia, 75*, 646–655.
- Krumm, S., Schmidt-Atzert, L., & Lipnevich, A. A. (2014). Specific cognitive abilities at work: A brief summary from two perspectives. *Journal of Personnel Psychology, 13*(3), 117–122.
- 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*(10), 1655–1677.

- Lang, J. W., Kersting, M., Hülshager, U. R., & Lang, J. (2010). General mental ability, narrower cognitive abilities, and job performance: The perspective of the nested - factors model of cognitive abilities. *Personnel Psychology, 63*(3), 595–640.
- Leutner, F., Codreanu, S. C., Brink, S., & Bitsakis, T. (2023). Game based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology, 13*, 942662.
- Li, B., & Bai, Y. (2020). Market signaling of educational background: Evidence from a field experiment. *Economic Research, 55*(10), 176–192.
- [李彬, 白岩. (2020). 学历的信号机制: 来自简历投递实验的证据. *经济研究, 55*(10), 176–192.]
- Lind, S. E., Bowler, D. M., & Raber, J. (2014). Spatial navigation, episodic memory, episodic future thinking, and theory of mind in children with autism spectrum disorder: Evidence for impairments in mental simulation? *Frontiers in Psychology, 5*, Article 1411.
- Li, Y., Yu, Z., Han, X., Liu, T. (2015). General Mental Ability Predicts Performance: The State and Mechanism. *Advances in Psychological Science, 23*(3), 448–459.
- [李英武, 于宙, 韩笑, 刘婷安. (2015). 一般心理能力(GMA)预测绩效: 现状、机制及趋势. *心理科学进展, 23*(3), 448–459.]
- Maruff, P., Thomas, E., Cysique, L., Brew, B., Collie, A., Snyder, P., & Pietrzak, R. H. (2009). Validity of the CogState brief battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Archives of Clinical Neuropsychology, 24*(2), 165–178.
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*(3), 272–277.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693–2698.

- Monahan, J. S., Harke, M. A., & Shelley, J. R. (2008). Computerizing the mental rotations test: Are gender differences maintained? *Behavior Research Methods*, *40*(2), 422–427.
- Morgeson, F. P., Delaney-Klinger, K., & Hemingway, M. A. (2005). The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *Journal of Applied Psychology*, *90*(2), 399–406.
- Mount, M. K., Oh, I. S., & Burns, M. (2008). Incremental validity of perceptual speed and accuracy over general mental ability. *Personnel Psychology*, *61*(1), 113–139.
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance*, *2*(3), 183–200.
- Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310.
- Nye, C. D., Ma, J., & Wee, S. (2022). Cognitive ability and job performance: Meta-analytic evidence for the validity of narrow cognitive abilities. *Journal of Business and Psychology*, *37*(6), 1119–1139.
- Pak, R., Czaja, S. J., Sharit, J., Rogers, W. A., & Fisk, A. D. (2008). The role of spatial abilities and age in performance in an auditory computer navigation task. *Computers in Human Behavior*, *24*(6), 3045–3051.
- Pan, D. N., Xie, H., Zeng, Y., Zhou, Y., Lin, C., Ma, X., ... & Xue, G. (2024). The development and validation of a tablet-based assessment battery of general cognitive ability. *BMC Psychology*, *12*(1), 778.
- Pedersen, M. K., Díaz, C. M. C., Wang, Q. J., Alba-Marrugo, M. A., Amidi, A., Basaiawmoit, R. V., Bergenholtz, C., Christiansen, M. H., Gajdacz, M., Hertwig, R., Ishkhanyan, B., Klyver, K., Ladegaard, N., Mathiasen, K., Parsons, C., Rafner, J., Villadsen, A. R., Wallentin, M., Zana, B., & Sherson, J. F. (2023). Measuring cognitive abilities in the wild: Validating a population - scale game - based cognitive assessment. *Cognitive Science*, *47*(6), e13308.
- Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, *39*(2), 313–338.
- 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.

- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2(4), 313–345.
- Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, 59, 419–450.
- Sackett, P. R., Zedeck, S., & Fogli, L. (1988). Relations between measures of typical and maximum job performance. *Journal of Applied Psychology*, 73(3), 482–486.
- Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, 22(6), 800–811.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53–61.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86(1), 162–173.
- Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.
- Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, 25(1), 12–27.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703.
- Sheppard, K. W., & Cheatham, C. L. (2017). Validating the electric maze task as a measure of planning. *Journal of Cognition and Development*, 18(2), 309–322.
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, 63, 106–117.
- Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *American Journal of Psychology*, 15(2), 201–293.

- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59*(4), 745–759.
- Sternberg, D. A., Ballard, K., Hardy, J. L., Katz, B., Doraiswamy, P. M., & Scanlon, M. (2013). The largest human cognitive performance dataset reveals insights into the effects of lifestyle factors and aging. *Frontiers in Human Neuroscience*, *7*, Article 292.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500–517.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29.
- Troll, E. S., Venz, L., Weitzenegger, F., & Loschelder, D. D. (2022). Working from home during the COVID - 19 crisis: How self - control strategies elucidate employees' job performance. *Applied Psychology: An International Review*, *71*(3), 853–880.
- van Aken, L., van der Heijden, P. T., Oomens, W., Kessels, R. P., & Egger, J. I. (2019). Predictive value of traditional measures of executive function on broad abilities of the cattell–horn–carroll theory of cognitive abilities. *Assessment*, *26*(7), 1375–1385.
- 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.
- Varriale, V., van der Molen, M. W., & De Pascalis, V. (2018). Mental rotation and fluid intelligence: A brain potential analysis. *Intelligence*, *69*, 146–157.
- Wang, E. H. J., Lai, F. H. Y., Leung, W. M., Shiu, T. Y., Wong, H., Tao, Y., Zhao, X., Zhang, T. Y. T., & Yee, B. K. (2025). Assessing rapid spatial working memory in community-living older adults in a virtual adaptation of the rodent water maze paradigm. *Behavioural Brain Research*, *476*, 1–15.
- Wilmot, M. P., & Ones, D. S. (2019). A century of research on conscientiousness at work. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *116*(46), 23004–23010.
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, *29*(1), 64–77.

- Sun, X., Li, J., & Fu, Z. (2018). Using game log-file to predict students' reasoning ability and mathematical achievement: An application of machine learning. *Acta Psychologica Sinica*, 50(7), 761–770.
- [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.]
- Xu, F., Wang, Q., Chen, M., Liang, X., & Yu, Y. (2019). Construction and application of employer value proposition in enterprise: A case study of S company. *Human Resources Development of China*, 36(3), 145–157.
- [许昉昉, 王青, 陈美伶, 梁骁, 喻依. (2019). 企业雇主价值主张的构建与应用: 基于 S 公司的案例研究. *中国人力资源开发*, 36(3), 145–157.]
- Xu, J., & Li, Z. (2021). Game-based psychological assessment. *Advances in Psychological Science*, 29(3), 394–403.
- [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.]
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, 20(1), 1–19.

From Recruitment to the Workplace: Differential Predictive Effects of Cognitive Abilities

ZOU Yuhan¹, GUAN Jian^{1,2}, CHEN Hao^{1,3}

(¹ Department of Social Psychology, School of Sociology, Nankai University, Tianjin 300350, China) (² Shanghai Haneo GBA

Intelligent Technology Co.,Ltd. Shanghai 200120, China) (³ Guangdong Provincial Key Laboratory of Adolescent Mental Health and

Psychological Development, Guangzhou 510275, China)

Abstract

In personnel selection, the phenomenon in which candidates perform well during the selection process but fail to match expectations in subsequent job performance suggests that selection assessments may fail to distinguish between cognitive functions important to perform well during the selection process and cognitive functions that are important for successful job performance in post. Against this backdrop, by introducing a "stage-matching" perspective to address this performance-job mismatch, we conducted two field studies using game-based assessments adapted from classic psychological experimental paradigms, aiming to reveal the relative predictive differences of specific cognitive abilities across real-world organizational stages.

Study 1 included 789 graduate applicants in a campus recruitment program at a Fortune Global 500 company. These participants completed a suite of Game-Based Assessments (GBA) specifically designed to measure specific cognitive abilities. The GBA battery comprised four tasks: the Corsi Block Task for working memory, the Mental Rotation Test for spatial ability, the Flanker Task for inhibitory control, and a Maze Task for planning and flexibility. Prior to the main analysis, psychometric evaluations were conducted, confirming the test-retest reliability, structural validity, and criterion-related validity of these game-based measures. Hiring outcomes were drawn from official organizational records. Study 2 tracked 446 newly hired employees from four different companies who had previously completed the GBA during their recruitment

and successfully passed the standard interview process. These participants were hired for diverse roles, including management trainees, marketing, and technical positions. Due to the heterogeneity of performance appraisal systems across different organizations and roles, performance metrics were categorized into "high performance" and "low performance" groups based on internal organizational standards to ensure comparability. Their job performance was evaluated by supervisors three to six months after onboarding. This design enabled a robust comparison of the predictive validity of specific cognitive abilities between recruitment and workplace contexts. In Study 1, the group of selected candidates exhibited significantly superior performance on most cognitive indicators compared to non-selected candidates, indicating a broad cognitive advantage among hired individuals. Since the selection stage represents a "maximal performance" context characterized by time pressure, high stakes, and novel information, specific cognitive resources are taxed differently. Logistic regression and XGBoost-SHAP analyses further confirmed working memory accuracy as a relatively stronger predictor of selection outcomes ($B = 0.97$, $OR = 2.64$, $p < 0.001$; $SHAP = 0.67$), demonstrating that rapid information updating is the relatively more critical factor for success in competitive recruitment settings. In Study 2, high-performance employees similarly outperformed the low-performance group on the majority of cognitive tasks. In contrast to the selection phase, the workplace represents a "typical performance" context characterized by persistence, multi-tasking, and frequent distractions. Regression analyses identified inhibitory control accuracy as a relatively higher predictor of job performance ($B = 0.58$, $OR = 1.78$, $p < 0.001$; $SHAP = 0.32$), while other cognitive measures contributed smaller, auxiliary effects, reflecting the essential role of resisting interference and maintaining goal-directed focus during routine professional tasks.

The findings of this study not only extend the ability–performance compatibility principle and refine the theory of stage-specific job performance, but also provide empirical support for the validity of cognitive ability assessments in predicting real working performance. These results offer organizations psychological insights for talent selection and management, suggesting that

targeted assessment of specific cognitive abilities may serve as a reference for human resource practices across different career stages.

Keywords recruitment, job performance, working memory, inhibitory control, game-based assessment

附录

附录 1

在项目初期，研究者与多家游戏开发公司进行了沟通，确保测评工具在科学性、可行性和适用性方面的平衡。开发过程中关注以下几个方面：界面设计简洁直观，减少认知负担；优化交互流程，确保易用性与快速上手；保证技术方案先进，数据采集稳定，确保行为数据的完整性和可分析性。游戏页面采用科技感风格，以星空背景提升沉浸体验，如正文图 1 所示。每个测试前提供引导提示，包括操作要求、测试时长(约 10-15 分钟)、设备使用说明以及测试顺序要求。每项测试都有明确目标和操作说明，确保用户理解流程，减少非能力性偏差。评估过程中未使用背景音乐，避免听觉干扰，但关键节点设置了实时音效反馈(如点击、拖拽提示)，以强化操作感知和情境体验。

工作记忆测试：每个信封将在画面中显示 2 秒，打开后呈现一个数字或简单算式，随后自动关闭。参与者需在信封全部关闭后 10 秒内，按数值大小与空间位置(从左到右、数值由小到大)将信封拖放至目标区域。为了减少操作不熟悉带来的误差，正式测试前设置了三信封的试玩阶段。正式测评包括 10 轮测试，每轮信封数量逐步递增(3 至 7 个)，此设计旨在避免过早超出个体能力范围，评估参与者在不同认知负荷下的信息处理表现。

空间表征测试：参与者需观察由立方体单元组成的不规则三维结构，并在规定时间内推断并绘制其主视图、俯视图和左视图。正式测评包含五个试次，测试任务复杂度逐步增加，先为 $3 \times 3 \times 3$ 的立方体，随后为两个 $4 \times 4 \times 4$ 和两个 $5 \times 5 \times 5$ 的立方体。每轮测试限时 60 秒，超时自动结束，测试难度逐步递增，以反映不同认知负荷下的表现。交互设计允许通过键盘与鼠标旋转立体图形，以支持多角度观察并减少视角偏差。所有结构均依据一致的随机规则生成，确保难度可比性和测评结果的可靠性。

抑制控制测试：本研究使用飞机、云朵和闪电等符号替代传统箭头，以提升沉浸感与参与度。测试任务包括四种规则情境：1) 中心与侧面飞机方向一致时，按中心飞机方向反应；2) 若方向不一致，则依据侧面飞机方向反应，诱发认知冲突；3) 中心为云朵时忽略干扰，仅反应中心飞机；4) 若出现闪电符号，则抑制当前反应，不按键。测试任务共 60 次试次，每种规则呈现 15 次，且前五次试次不包含闪电抑制条件。每次刺激呈现 500 毫秒，参

与者需在此限时内完成反应。系统对反应正确性与速度提供即时反馈，提示“过快”或“过慢”时，反馈持续 300 毫秒。

路径规划测试：迷宫为 10×10 方格地图，参与者需从起点出发，尽量以更少步数到达终点，完成 5 轮测试。每轮测试限时 5 分钟，步数定义为沿某一方向行进至停止的单位。为了增加挑战性，地图中嵌入红花、绿地和石头三类标识，分别对应惩罚、奖励与障碍。为控制测试难度，地图复杂度统一，系统随机分配地图，确保测试在认知要求上的可比性，防止记忆化或简化测试，提升测评结果的可靠性与有效性。

附录 2

本研究首先选取国家公务员行政职业能力倾向测验中的数理逻辑部分作为效标工具。该测验主要指向智力因素(辛自强, 2018), 内容涵盖数字推理(如等差、自然数列)、数量运算(如排列组合、几何运算)以及图表分析三个维度。共 592 名参与者同时完成了该测验与 GBA 测评, 各指标的描述性统计与相关系数矩阵结果见表 S1。

表 S1 描述性统计与相关系数矩阵

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1 数理逻辑分数	13.65	4.84	—							
2 工作记忆测试 正确数	31.48	6.08	0.64***	—						
3 空间表征测试 正确数	9.26	2.68	0.57***	0.44***	—					
4 抑制控制测试 正确数	40.30	9.21	0.58***	0.46***	0.50***	—				
5 路径规划测试 最优步数差	4.20	1.39	-0.48***	-0.37***	-0.38***	-0.36***	—			
6 工作记忆测试 作答时长	81020.4	10089.0	-0.09*	-0.07	-0.14***	-0.07	0.05	—		
7 空间表征测试 作答时长	228909.3	27386.0	-0.27***	-0.21***	-0.32***	-0.28***	0.15***	0.30***	—	
8 抑制控制测试 作答时长	70182.4	4962.2	-0.00	0.03	-0.06	-0.00	0.07	0.08	0.13**	—
9 路径规划测试 作答时长	134193.6	54303.3	0.17***	0.17***	0.13**	0.07	-0.20***	0.09*	0.07	0.01

注: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 下同。

分析结果显示, GBA 测评中的工作记忆测试(第一关)、空间表征测试(第二关)及抑制控制测试(第三关)的正确数与数理逻辑总分的相关系数分别为 0.64、0.57 与 0.58($p < 0.001$); 路径规划测试(第四关)的最优步数差与数理逻辑总分的相关系数为-0.48($p < 0.001$), 即步数偏差越小, 数理逻辑得分越高。基于 Schmid-Leiman 正交转换(1957)提取的 g 因子与数理逻辑总分呈显著正相关($r = 0.57, p < 0.001$)。

研究者另招募了 293 名参与者, 同时完成 GBA 测评和《工作记忆量表》(Vallat-Azouvi et al., 2012)。该量表共 30 个条目, 以较高生态效度著称。尽管名为“工作记忆”, 该量表

实质上综合评估了短时记忆(也涉及心算与书面文本理解)、注意力(如克服分心、双任务处理)以及执行能力(如决策、规划和任务转换)。鉴于上述构念间存在重叠,且已有研究指出该量表总分与视觉广度、瑞文矩阵等一般认知测验显著相关,本研究将其总分作为反映个体整体执行功能水平的效标,各指标的描述性统计与相关系数矩阵结果见表 S2。

表 S2 描述性统计与相关系数矩阵

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1 工作记忆量表 分数	119.75	17.88	—							
2 工作记忆测试 正确数	32.27	6.66	0.38***	—						
3 空间表征测试 正确数	8.61	2.63	0.37***	0.44***	—					
4 抑制控制测试 正确数	53.15	9.73	0.40***	0.44***	0.36***	—				
5 路径规划测试 最优步数差	3.89	1.30	-0.44***	-0.40***	-0.41***	-0.39***	—			
6 工作记忆测试 作答时长	259054.6	36964.0	-0.08	-0.16**	-0.05	-0.04	0.02	—		
7 空间表征测试 作答时长	283953.9	33497.7	-0.00	-0.13*	-0.15*	-0.07	0.01	0.28***	—	
8 抑制控制测试 作答时长	111621.2	3753.1	0.25***	0.26***	0.20***	0.63***	-0.22***	-0.03	-0.05	—
9 路径规划测试 作答时长	129989.8	53556.1	0.24***	0.14*	0.16**	0.14*	-0.29***	0.02	0.15**	0.16**

分析显示, GBA 各关卡的正确数与量表总分呈显著正相关($r_s = 0.37 \sim 0.40, p_s < 0.001$), 最优步数差与量表总分呈显著负相关($r = -0.44, p < 0.001$)。此外, 基于 Schmid-Leiman 正交转换(1957)提取的 g 因子与量表总分的相关系数最高($r = 0.52, p < 0.001$)。

最后, 研究者招募了 349 名参与者, 同时完成 GBA 测评与《客体-空间表象与言语认知风格问卷》中的空间表象分量表(Blazhenkova & Kozhevnikov, 2009)。该分量表包含 15 个条目, 是评估个体视觉空间能力的有效工具(鲍旭辉 等, 2012), 故被选作衡量空间表征水平的效标。各指标的描述性统计与相关系数矩阵结果见表 S3。

表 S3 描述性统计与相关系数矩阵

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1 空间表象量表 分数	51.98	9.01	—							
2 工作记忆测试 正确数	31.64	6.80	0.26***	—						
3 空间表征测试 正确数	8.90	2.80	0.49***	0.40***	—					
4 抑制控制测试 正确数	53.40	9.71	0.27***	0.36***	0.29***	—				
5 路径规划测试 最优步数差	3.92	1.36	-0.37***	-0.38***	-0.33***	-0.29***	—			
6 工作记忆测试 作答时长	266871.1	37038.9	0.00	-0.17**	-0.03	0.06	0.04	—		
7 空间表征测试 作答时长	286131.8	37275.2	-0.05	-0.11*	-0.17**	-0.02	0.03	0.37***	—	
8 抑制控制测试 作答时长	111613.2	3119.6	-0.01	-0.01	-0.10	0.00	0.03	0.12*	0.03	—
9 路径规划测试 作答时长	135702.0	59281.1	0.02	0.16**	0.07	0.08	-0.11*	0.09	0.16**	0.01

分析结果显示, GBA 第二关(空间表征关卡)的正确数与该效标总分的相关性最高($r = 0.49, p < 0.001$)。相比之下, 基于 Schmid-Leiman 正交转换(1957)提取的 g 因子及其他关卡正确数虽然也与效标呈显著正相关($r_s = 0.11 \sim 0.27, p_s < 0.05$), 但相关强度较弱; 最优步数差与效标呈显著负相关($r = -0.37, p < 0.001$)。

综上, 研究者对本研究中所使用的测评工具进行了三组效度检验, 上述结果支持 GBA 测评具有良好的效标关联效度。

附录 3

在研究 2 中,研究者采用与研究 1 相同的结构效度检验方法。Bartlett 球形检验结果($\chi^2 = 323.94, p < .001$)和 KMO 检验结果(0.64)表明数据适合进行因子分析。在此基础上,首先对四项测试构建验证性因子分析模型。与既有游戏化测评研究的结果(Landers et al., 2022)以及研究 1 的结果一致,游戏测试的关卡情境与行为诱发的一致性导致测验指标之间存在载荷交叉,该模型的整体拟合度不理想($\chi^2(20) = 113.77, p < .001, CFI = 0.69, RMSEA = 0.10$)。与研究 1 相同,参照 Landers et al. (2022)关于游戏化测评结构效度检验的处理建议,在允许部分测试指标残差相关的条件下对模型进行重新拟合后,模型拟合度得到一定程度改善($\chi^2(17) = 52.99, p < .001, CFI = 0.88, RMSEA = 0.07$)。

为进一步检验测评是否同时反映一般认知能力与特定认知能力,本研究采用 Schmid-Leiman(1957)方法的正交层级模型探索测评结构。结果显示,在基于四类认知测试构建的正交层级模型中,g 因子作为第一因子,解释共同方差占比为 37.28%。与此同时,各测试在其对应的下位因子上仍保留载荷,支持该测评不仅能够反映一般认知能力,也包含特定认知能力的成分,为后续检验它们在不同情境下的差异化预测作用提供了结构基础。以上结果表明,研究 2 所使用的 GBA 结构效度总体可接受。

参考文献(附录)

- Bao, X., He, L., Shi, M., & You, X. (2012). The Model and Measurement of Object-Spatial Image and Verbal Cognitive Style. *Advances in Psychological Science*, 20(4), 523–531.
[鲍旭辉, 何立国, 石梅, 游旭群. (2012). 客体-空间表象和言语认知风格模型及其测量. *心理科学进展*, 20(4), 523–531.]
- Xin, Z. (2018). Psychological issues inside social governance. *Advances in Psychological Science*, 26(1), 1–13.
[辛自强. (2018). 社会治理中的心理学问题. *心理科学进展*, 26(1), 1–13.]
- Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology*, 23(5), 638–663.

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*(10), 1655–1677.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.

Vallat-Azouvi, C., Pradat-Diehl, P., & Azouvi, P. (2012). The Working Memory Questionnaire: A scale to assess everyday life problems related to deficits of working memory in brain injured patients. *Neuropsychological Rehabilitation, 22*(4), 634–649.