1. 研究系统比较了显性测量（自我报告）与隐性测量（行为任务）在积极自我认知中的结构特性与功能预测效度。通过相关分析、因子分析、网络建模与聚类检验，我们发现两类测量在统计结构上高度分离：自我报告变量呈现出良好的结构聚合与一致性，而行为任务变量结构松散、内部关联弱。进一步利用线性（LASSO）与非线性（随机森林）回归模型对四项心理健康指标进行预测分析，结果显示：显性测量虽然结构清晰，但预测力有限，LASSO 模型中 R² 接近 0；而结构松散的隐性测量在随机森林模型中展现出稳定且显著的预测优势（R² 可达 0.78）。这一对比揭示了典型的“结构—功能悖论”：即结构效度较高的测量方式未必具备更强的功能预测力，反之，结构弱的隐性指标可能蕴含非线性、高阶交互等复杂的预测信息。本研究从功能性与结构性两方面，讨论了测量的结构与实用价值，强调了测量评估中功能性指标的重要性；“结构—功能悖论”也表明，个体的自我认知不是一个静态结构，而是一个具有动态加工、情境敏感与多系统协同的复杂心理系统。
2. 研究思路

1.研究目的

本研究旨在探讨显性测量（如自我报告问卷）与隐性测量（如行为任务）在积极自我认知中的结构特性及其对心理健康变量的预测效度，检验两者是否存在“结构—功能悖论”现象。

2.数据结构与变量

经过共性筛选后，排除了项目间共性较低的条目。研究样本保留 133 个自我报告条目，7 个行为任务指标，以及四项心理健康变量（自尊、抑郁、焦虑、主观幸福感）。

3.分析步骤

（1）结构分析：计算变量相关矩阵、十折交叉验证与网络建模，评估两类测量的结构耦合性；

（2）结构比较：通过聚类与 Mantel 检验进一步检验测量结构系统的相似性；

（3）功能预测分析：构建 LASSO 与随机森林模型，分别比较显性与隐性变量对心理健康变量的预测效度；

（4）公平性控制：为避免条目数量差异带来的偏差，对显性测量采用等量抽样+重复建模策略。

4.核心发现

（1）显性测量结构清晰但预测不强，隐性测量结构弱但功能预测力高；

（2）非线性模型（如随机森林）在处理复杂、高维度心理变量时表现更优，特别是对行为任务中的非线性结构具有更强解释力。

5.研究意义

本研究强调：心理测量应超越对结构清晰性的过度追求，重视预测功能与个体适应的实用价值。显性与隐性测量的差异不只是技术层面的问题，而是反映了不同心理加工系统的认知机制和功能表现。

1. 数据处理

1.内隐联结效应IAT的测量与计算

本研究采用内隐联结测验（Implicit Association Test, IAT）评估个体在两个领域中的积极自我认知偏好，并分别为道德领域与能力领域。IAT 任务旨在捕捉个体对“自我/他人”与“积极/消极”词汇之间的自动联结强度。通过反应时数据，计算标准化的 D 值（D-score）作为内隐偏好的效应量指标。

原始反应时数据中包含四个关键 block（A1, A2, B1, B2），其中 A1 与 A2 表示兼容条件（如“我-积极/他人-消极”配对），B1 与 B2 表示不兼容条件（如“我-消极/他人-积极”配对）。在试次选择方面，我们选择保留 IAT 中关键测量阶段（screen\_id = 3, 4, 6, 7）中的反应数据；在反应式剔除方面，我们剔除所有反应时间超过 10,000 毫秒的异常试次；为消除错误反应对结果的干扰，参照 Greenwald 等人（2003）推荐的方法，对所有错误反应试次进行校正：将错误试次的反应时替换为该 block 内的平均反应时加上 600 毫秒惩罚值。之后，对每位被试在每个 block 下的反应时数据进行汇总，计算每个条件的平均反应时、标准差和试次数。

最终，我们按照标准 D 值计算方法进行效应量提取。分别计算 B1 与 A1、B2 与 A2 之间的平均反应时差，并将其除以对应条件下的合并标准差（pooled SD）进行标准化，获得每位被试在moral领域和ability领域下的 IAT D 值。根据 D 值的绝对大小，将效应划分为三类：|D| ≤ 0.2 为小效应，0.2 < |D| ≤ 0.5 为中等效应，|D| > 0.5 为大效应。

2.ALT实验范式与指标计算方法

在本研究中，我们使用了 ALT（Associative Learning Task）实验范式来测量参与者在道德与能力两个维度下，对“自我”与“朋友”关联积极特质的加工差异。该任务通过呈现一系列“匹配”（如“积极词汇-自我”）与“不匹配”（如“积极词汇-朋友”）的组合，要求被试做出快速反应，从而评估其内隐的积极自我关联倾向。

数据清洗方面，首先筛选了所有来自正式实验阶段的试次并排除反应时小于200ms或大于1200ms 的异常值，以保证数据质量。此外，对于缺失的“正确性”标记（correct），默认按错误处理，防止遗漏。同时，我们提取了两个维度的关键指标：反应时差异（ALT\_SE\_rt） 与信号检测敏感性指数（ALT\_SE\_d′），分别反映个体对匹配刺激的加工速度和辨别能力。

具体计算如下：

（1）ALT\_SE\_rt 指标计算：

首先选取所有“匹配”类型的试次（identit为match），并筛选其中“正性词汇”条件（valence = positive）。对于每位参与者，在“自我”与“朋友”条件下分别计算平均反应时（avg\_rt）。用 friend 条件减去 self 条件的平均反应时，得出 ALT\_SE\_rt 值。该值越大，代表对“自我-积极”匹配的加工越快，表征更强的积极自我加工倾向。

（2）ALT\_SE\_d′ 指标计算：

使用信号检测理论（Signal Detection Theory）框架，对每位参与者在“正性词汇”条件下进行分类统计。将试次按反应正确性与刺激类型（match vs. nonmatch）划分为：击中（hit）、错过（miss）、误报（false alarm） 与 正确拒绝（correct rejection）。计算击中率（hit rate） 与 误报率（false alarm rate），并采用调整策略防止极端值（如 hit rate=1 或 FA rate=0）导致Z值计算错误。根据公式，计算每位被试在“自我”与“朋友”条件下的 d′ 值。最终用 self 条件下的 d′ 减去 friend 条件下的 d′，得出 ALT\_SE\_d 值。该值越大，表示被试对“自我-积极”刺激辨别更敏感，具备更高的积极自我识别能力。

最终得到每位被试在 moral 与 ability 两个维度下的 ALT\_SE\_rt 与 ALT\_SE\_d 指标，分别储存在 moral\_ALT\_rt、ability\_ALT\_rt 及 moral\_ALT\_d、ability\_ALT\_d 四个变量中，用于数据分析。

3.3. SRET实验范式与指标计算方法

在本研究中，我们使用了 SRET（Self-Referent Encoding Task）范式及其延伸任务，系统测量参与者在“道德”与“能力”两个维度下的积极自我认知水平。任务包括三种子任务：词汇判断任务（SRET-EW 与 SRET-RT）、新旧词判断任务（SRET-RJ1）与来源判断任务（SRET-RJ2）。每个任务分别从评估反应偏好、加工速度与记忆敏感性角度，提取了多个自我相关指标。

数据清洗方面，首先仅保留正式实验阶段的试次，并剔除所有练习或非正式试次。在每类任务中，统一剔除 8 个用于练习或与研究目标不相关的干扰词汇。此外，对于反应时低于200毫秒或高于1200毫秒的试次，亦予以排除，以控制极端值对平均指标的干扰。缺失的判断正确性（correct）变量则默认标为错误，以避免漏判。

本研究共提取了八个 SRET 指标，分别对应三类任务，具体如下：

（1）SRET\_EW 指标计算

SRET\_EW 基于词汇判断任务数据。在每位被试的“自我-积极”与“自我-消极”条件中，分别统计其作出“是”判断的次数，两者之差表示自我条件下选择积极词汇的倾向。该指标在道德（morality）与能力（ability）两个领域分别计算，生成 moral\_SRET\_EW 与 ability\_SRET\_EW 两个变量。数值越大，表明被试更倾向于认同与自我相关的积极特质，反映其积极自我认知水平。

（2）SRET\_rt 指标计算

SRET\_rt 同样来源于词汇判断任务，但关注的是反应时数据。在每位被试“自我-积极”与“自我-消极”条件下，计算反应时均值之差。差值越大表示参与者对积极词汇加工更快，反映更高的积极自我加工效率。该指标同样在两个领域上独立计算，得到 moral\_SRET\_rt 与 ability\_SRET\_rt 两个变量。

1. SRET\_RJ1\_d 指标计算

该指标基于新旧词判断任务。任务要求参与者判断屏幕上呈现的词汇是否曾在之前出现，并作出“新”“旧”或“熟悉”反应。根据信号检测理论，将旧词被正确认定为“旧/熟悉”视为击中（hit），将新词误判为“旧/熟悉”为误报（false alarm）。通过公式 d′计算两者d′值 。以“自我”条件下的 d′ 减去“朋友”条件下的 d′ 值，生成 SRET\_RJ1\_d 指标，分别计算 moral\_SRET\_RJ1\_d 与 ability\_SRET\_RJ1\_d。

（4）SRET\_RJ2\_d 指标计算

该指标来源于来源判断任务。参与者需判断词汇来源于“自我”或“朋友”。以“自我”词汇为信号、朋友词汇为噪音，根据击中（hit）与误报（fa）计算 d′ 值。得出 SRET\_RJ2\_d 指标，分别为 moral\_SRET\_RJ2\_d 与 ability\_SRET\_RJ2\_d。

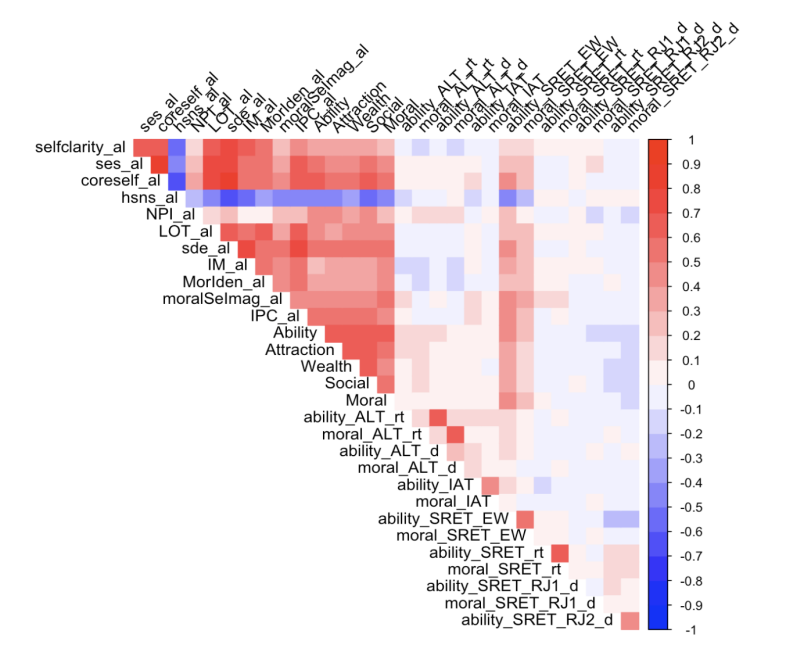
综上，研究共构建了八个SRET指标，覆盖加工偏好（SRET\_EW）、加工速度（SRET\_rt）与记忆敏感性（SRET\_RJ1\_d 与 SRET\_RJ2\_d）三个层面，分别在道德与能力两个维度上进行计算，反映个体在多个层面上的积极自我认知能力。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **指标名称** | **来源任务** | **计算方式** | **心理学含义** | **领域区分** |
| moral\_IAT | IAT任务 | D值（相容-不相容）反应时差/标准差 | 道德领域中积极自我联结的内隐强度 | Moral |
| ability\_IAT | IAT任务 | D值（相容-不相容）反应时差/标准差 | 能力领域中积极自我联结的内隐强度 | Ability |
| moral\_ALT\_rt | ALT任务 | friend - self 平均反应时差 | 对“自我-积极”匹配加工速度更快 | Moral |
| ability\_ALT\_rt | ALT任务 | friend - self 平均反应时差 | 对“自我-积极”匹配加工速度更快 | Ability |
| moral\_ALT\_d | ALT任务 | d prime = Z(hit) - Z(FA)，self - friend 差值 | 对“自我-积极”匹配的敏感性 | Moral |
| ability\_ALT\_d | ALT任务 | d prime = Z(hit) - Z(FA)，self - friend 差值 | 对“自我-积极”匹配的敏感性 | Ability |
| moral\_SRET\_EW | SRET任务 | positive\_self - negative\_self（yes响应数量差值） | 积极词更倾向归于自我 | Moral |
| ability\_SRET\_EW | SRET任务 | positive\_self - negative\_self（yes响应数量差值） | 积极词更倾向归于自我 | Ability |
| moral\_SRET\_rt | SRET任务 | negative\_self - positive\_self（反应时差值） | 反应更快地接受积极词归于自我 | Moral |
| ability\_SRET\_rt | SRET任务 | negative\_self - positive\_self（反应时差值） | 反应更快地接受积极词归于自我 | Ability |
| moral\_SRET\_RJ1\_d | SRET-新旧词判断任务 | d prime= Z(hit) - Z(FA)，self - friend 差值 | 判断“是否出现过”的敏感性，self为目标 | Moral |
| ability\_SRET\_RJ1\_d | SRET-新旧词判断任务 | d prime= Z(hit) - Z(FA)，self - friend 差值 | 判断“是否出现过”的敏感性，self为目标 | Ability |
| moral\_SRET\_RJ2\_d | SRET-来源记忆任务 | d prime= Z(hit) - Z(FA)，self为信号、friend为噪音 | 判断“来源是否是自己”的能力 | Moral |
| ability\_SRET\_RJ2\_d | SRET-来源记忆任务 | d prime= Z(hit) - Z(FA)，self为信号、friend为噪音 | 判断“来源是否是自己”的能力 | Ability |

三、相关分析——揭示隐性测量与显性测量的关系

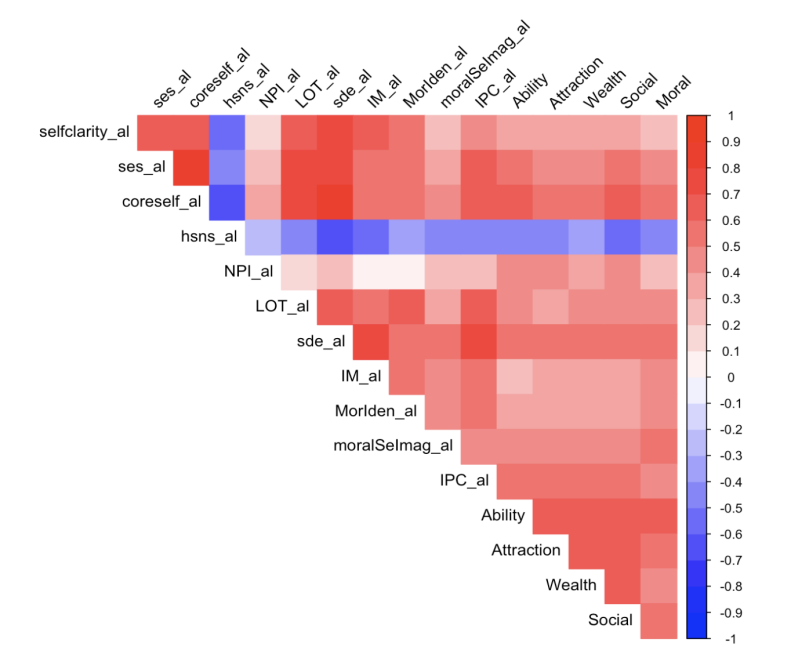
1. 对提取到的所有总分进行相关分析

为探究自评与认知任务指标之间的测量一致性，本研究计算了变量间的皮尔逊相关系数并使用 corrplot 绘制相关热图。结果显示，自我报告变量之间相关较高，而行为任务之间及其与自我报告的相关性普遍较低，提示显性与隐性测量结构存在分离。



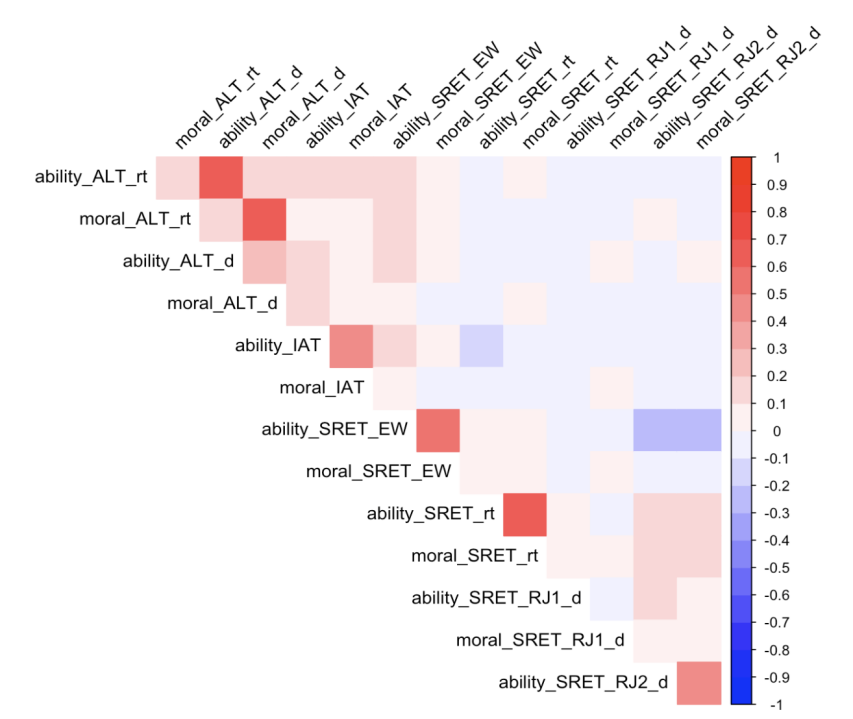
1. 对问卷总分进行相关分析

为检验各自我报告变量之间的关系，计算了前16列变量（自我报告）的皮尔逊相关系数并绘制相关热图。结果显示，大部分指标间存在中等到较高的正相关，反映出积极自我相关特质量表之间具有较强的一致性和结构耦合性。



1. 对行为任务进行相关分析

为检验各行为任务指标之间的关系，计算了第 17 至 30 列变量（行为任务）的皮尔逊相关系数，并绘制相关热图。结果显示，不同任务间的相关性整体偏低，部分指标甚至呈现无关或负相关，表明行为任务在内部结构上较为分散，可能反映了不同的认知加工机制或任务特异性较强。



1. 十折交叉验证的线性模型

在完成第一步的相关分析之后，为更深入理解积极自我认知各维度间的结构关系与变量间的预测潜力，本部分研究进一步采用了基于交叉验证的多元回归建模策略，以补充相关性分析中所揭示的表面关联，探索变量之间的统计整合性与个体差异中的系统性结构。

研究数据包括 503 位被试在 149 个自评条目上的完整得分，这些条目覆盖了积极自我认知的多个核心维度。尽管相关分析揭示了变量间的显著联系，但变量之间的预测关系与整体结构嵌套性仍需进一步探讨。因此，我们引入交叉验证建模方法，试图回答如下问题：在控制测量误差的前提下，一个积极自我认知变量在多大程度上可以由其余相关变量稳定预测？

本步骤采用嵌套式建模策略，将变量轮流作为因变量建模，同时在每次建模中使用十折交叉验证（10-fold cross-validation）评估模型的稳定性。整体流程如下：

1. 外层：轮流将变量作为预测目标

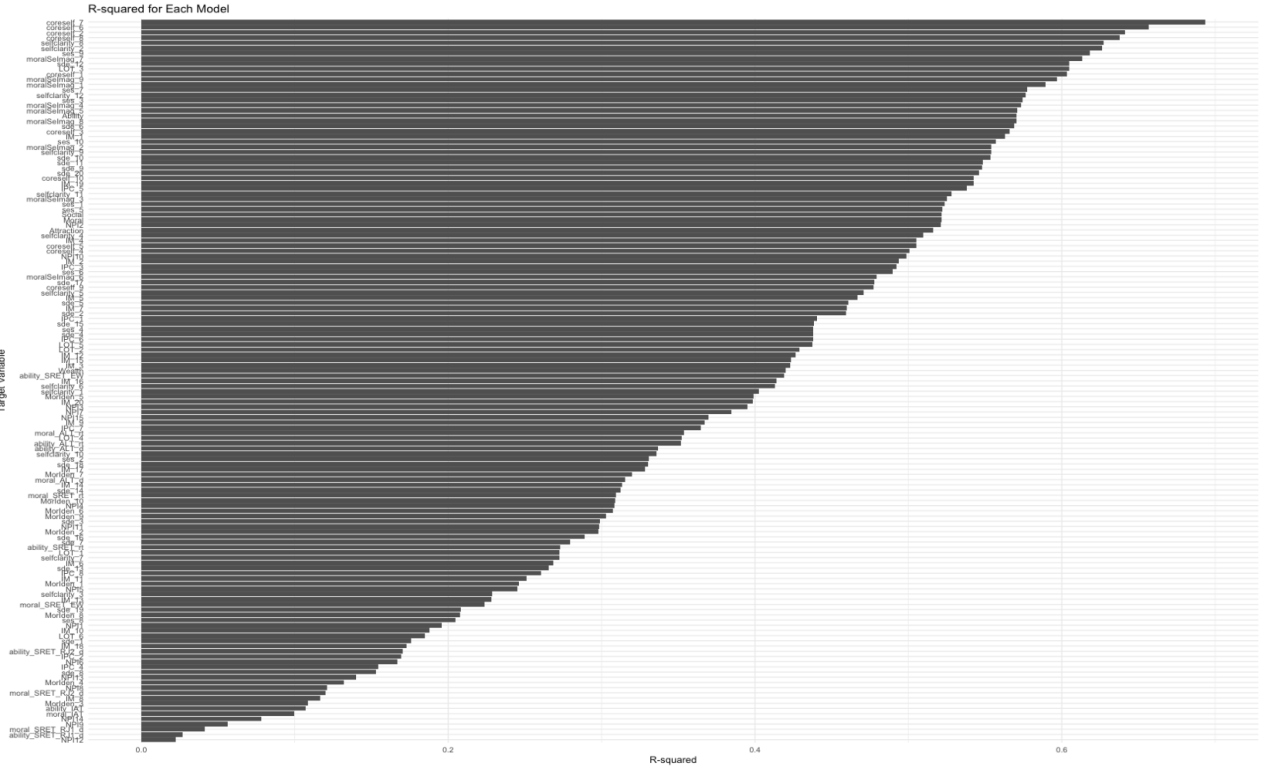
在外层循环中，依次将每一个变量设定为目标变量，使用其余 148 个变量作为自变量构建多元线性回归模型。此过程共运行 149 次，确保每个变量都作为预测目标被完整建模一次。

2. 内层：嵌套十折交叉验证

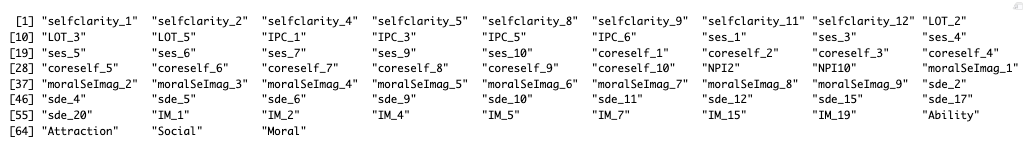
每一轮模型构建中，进一步执行十折交叉验证以确保结果的稳健性：首先将 503 个样本划分为 10 个fold，每轮以其中 9 个fold为训练集，剩余1个fold为验证集，训练模型并预测验证集。在此基础上重复10次，轮流更换验证集。最后，将 10 次验证的性能指标（RMSE、R²、MAE）进行平均，用于评估模型对该变量的整体预测力。

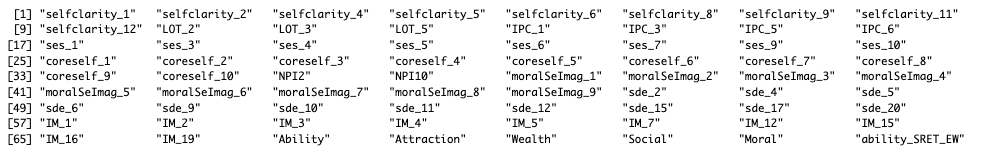
为识别变量间结构一致性较强的核心条目，本研究基于交叉验证线性模型计算了每个变量的预测性能。通过设定 R²(>0.3)、RMSE(<0.8) 和 R²(>0.4) 的筛选阈值，前者保留了66个条目，后者保留了73个条目。

从**统计结构角度**来看，自我报告变量之间呈现较强的相关性和结构耦合，导致它们更容易在建模中相互预测，而行为任务变量则表现出较低的结构协同性，难以融入整体预测模型框架。从**测量属性角度**，自我报告反映的是个体对自我认知的稳定表征，属于高结构性、低噪声的主观陈述类型；而行为任务则更依赖于任务情境和反应过程，结果波动性大、通用性弱，难以用其他变量稳定建模。从**心理学机制角度**，自我报告反映的是个体对自我特质的显性认知和长期评价，而行为任务更多激活的是即时加工或自动化反应，其结果未必直接映射到个体自我结构的认知层面，这也说明两类测量可能反映的是不同心理系统或认知路径中的加工阶段



保留条目：



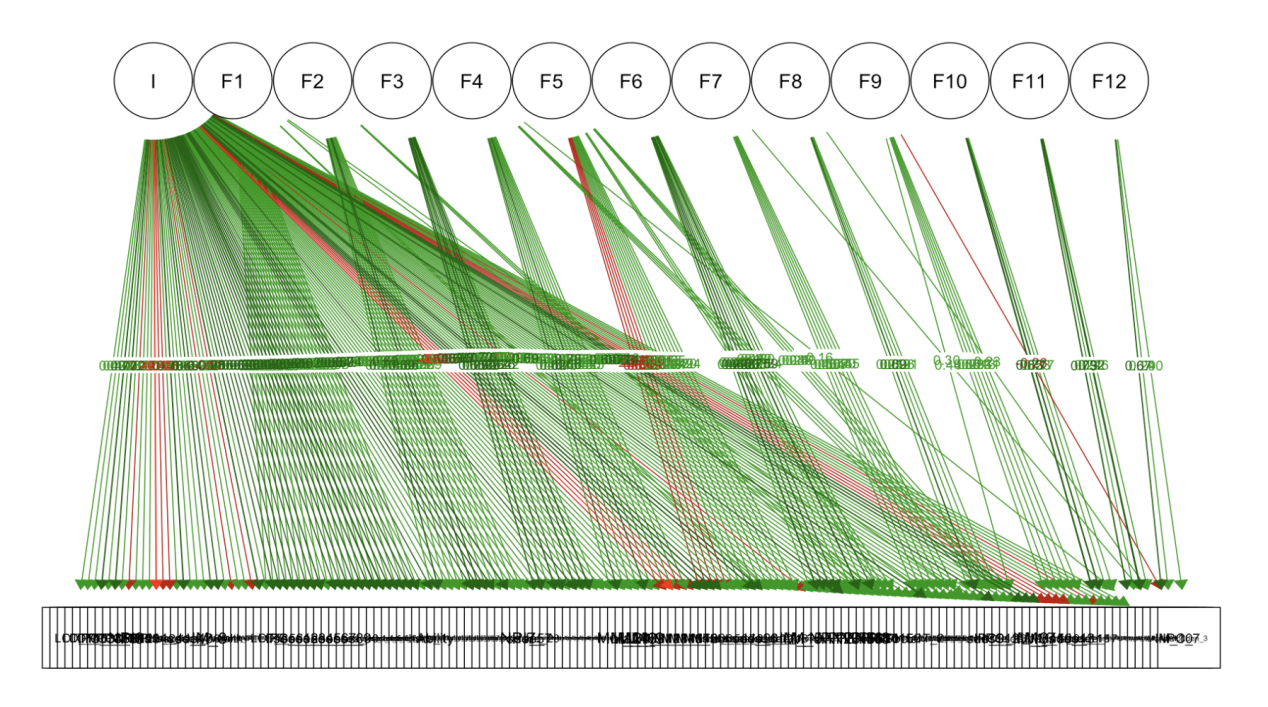


1. Bifactor Model分析

在进一步检验积极自我认知结构的潜在共性时，我们基于标准化数据构建了包含通用因子（general factor）与多个特异因子（specific factors）的双因子（bifactor）模型。拟合结果显示，模型的 RMSEA = 0.049 与 SRMR = 0.069，均处于可接受甚至理想范围，表明残差结构尚可；但 CFI = 0.735 与 TLI = 0.728 明显低于理想拟合标准（通常要求 > 0.90），提示该模型在整体结构拟合方面存在偏差。

结合模型结构与加载结果可见，尽管我们设定了一个统摄全部测量指标的通用因子（I），但其解释力相对有限，模型更倾向于表现为由多个特异因子主导的多维结构。这一结果暗示，在积极自我认知的测量中，尽管不同任务（如IAT、ALT、SRET）均意在揭示个体对自我持有的积极加工倾向，但它们可能反映了不同的心理机制——如反应偏好、加工速度、信号辨别能力等。也就是说，积极自我认知可能并非源于统一的潜在构念，而更可能是一种多维且任务特异性较强的结构。

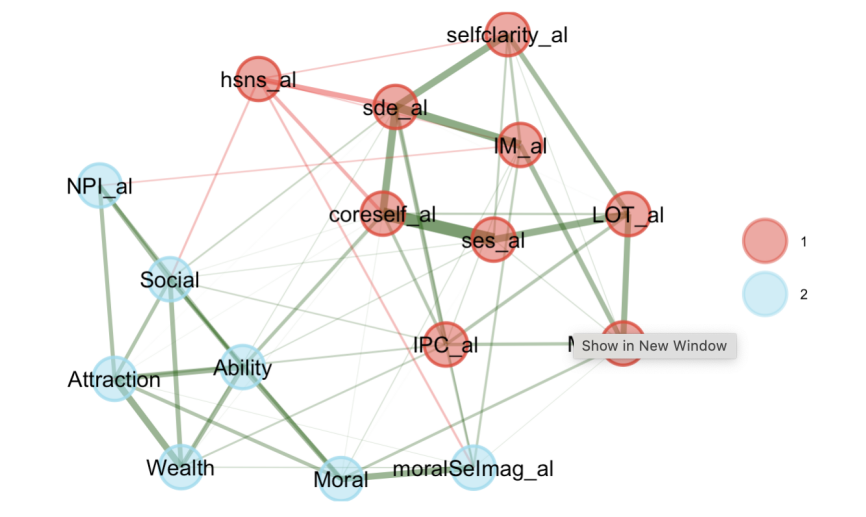
因此，本研究结果不支持将积极自我认知简化为单一共性因子所主导的结构模型，而支持其呈现出多种潜在加工成分的可能性。这也为未来在建构积极自我认知的理论模型时提供了证据支持：应充分考虑其心理加工的多样性与测量工具的异质性，而非假设其具有单一、内隐一致的潜在因子结构。



⸻

五、自我报告与行为任务的网络分析

为检验自我报告与行为任务变量的结构联结模式，分别对两类变量进行了基于图模型的网络分析（EGA, model = “glasso”），并构建了合并的邻接矩阵进行联合可视化。结果显示，两组变量在图结构上各自形成相对独立的子网络，节点之间的连接主要集中在各自组内，跨组边稀疏，表明自我报告与行为任务之间在结构层面呈现出明显的分离状态。



**综合第一部分研究，我们可以得出：**研究通过相关性分析、交叉验证预测建模、网络结构建模等多种方法，系统比较了自我报告与行为任务两类测量方式在积极自我认知测量中的表现。分析结果从多个层面共同揭示出两者在测量结构上的显著分离性。

1. 从统计结构角度

自我报告变量在多个分析中展现出高度的内部一致性和结构耦合，而行为任务指标则呈现出结构松散、内部关联弱、预测性低的特征。这表明两类变量在统计层面构成了两个相对独立的结构系统。

2. 从测量属性角度

自我报告测量的是个体对自我认知的稳定、可言说的主观评价，具有较高的一致性与结构化倾向；而行为任务则捕捉到个体在具体任务中的即时反应，受情境、任务设计和加工过程的影响较大，测量信噪比较低，难以在多个任务之间体现统一维度。

3. 从心理学机制角度

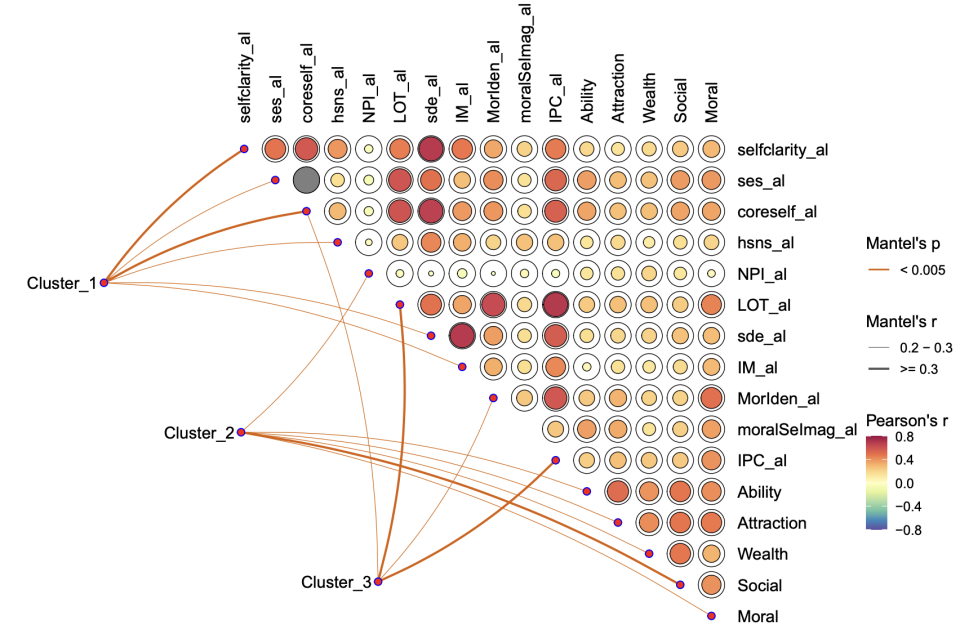
自我报告反映的是个体的长期自我表征与价值判断，而行为任务可能更多激活的是较为自动化或情境驱动的认知偏好与反应倾向。两类测量手段或许对应了不同心理加工系统，或不同层级的自我信息处理过程，因而在表现上缺乏显著一致性。

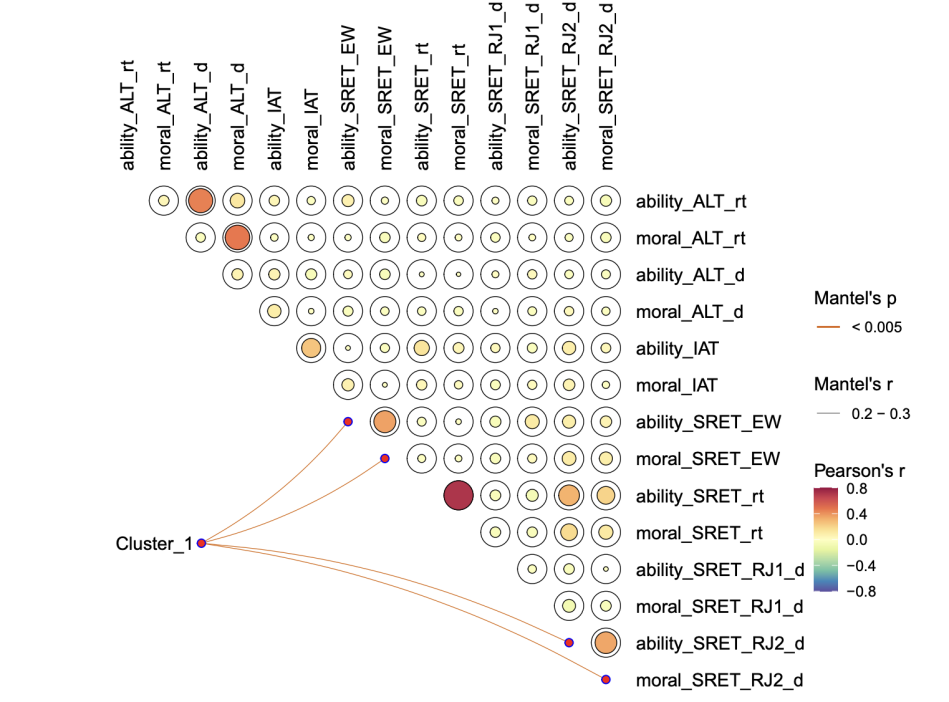
六、聚类分析与 Mantel 检验

为进一步理解问卷与行为任务之间相关性较低的现象是否源于其内部结构层级差异，本研究在初步相关分析的基础上，使用聚类分析与 Mantel 检验方法对两类测量的内部结构展开检验。具体而言，分别以问卷得分与行为任务指标为单位构建距离矩阵，并对各自变量进行层次聚类分析，观察是否存在明显的聚合模式。

由于 Mantel 检验对负值敏感，分析前对所有变量数据进行了平方处理。随后对变量之间的聚类结果与它们的相关性矩阵进行比对，使用 mantel\_test() 检验各聚类结构与变量相关性的耦合程度。我们将变量聚为三类，并筛选出相关性系数 r ≥ 0.2 且显著性 p < 0.005 的显著连线进行可视化。

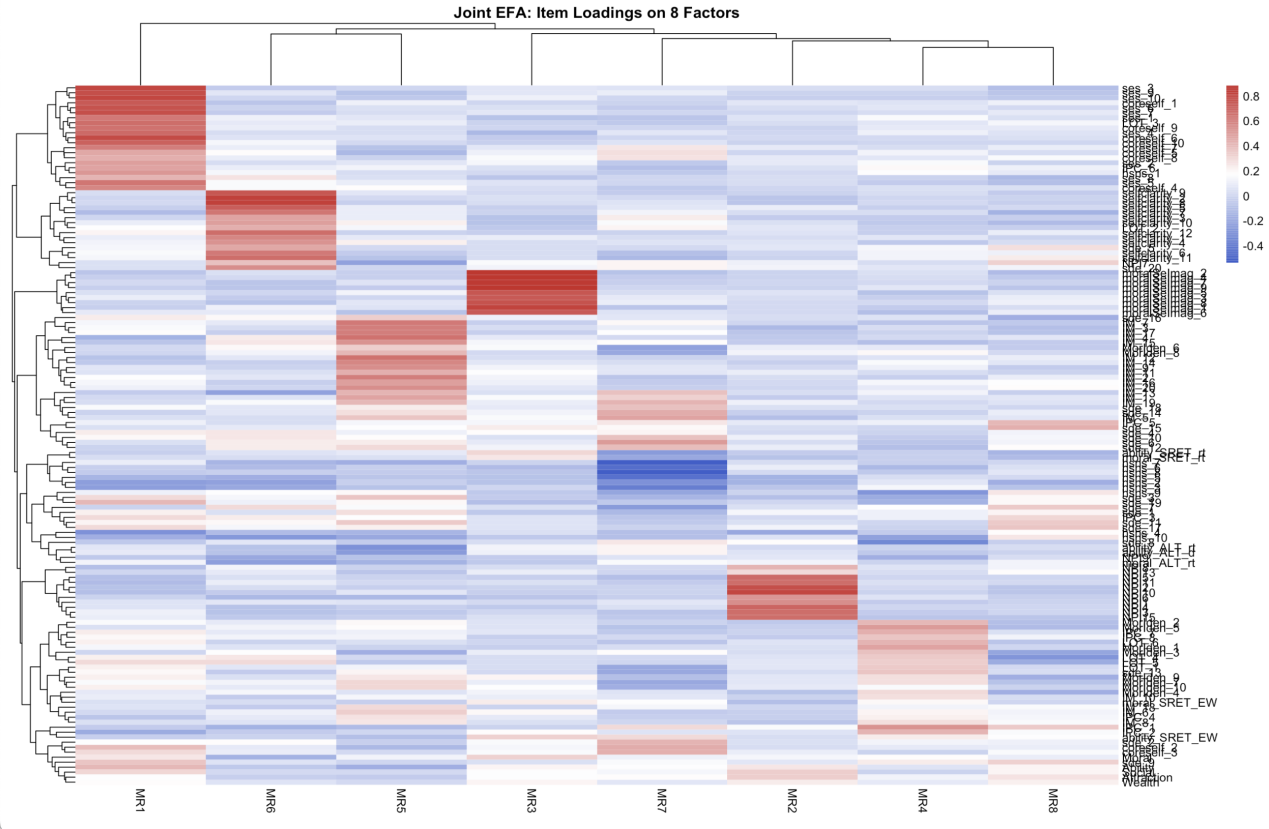
结果显示，问卷变量之间聚类结构清晰，且多个类簇间呈现显著的结构耦合性，表明问卷量表间存在相对稳定、分化明确的内部结构维度。而行为任务变量之间则缺乏清晰的聚类结构，Mantel 检验结果也未呈现显著耦合，暗示任务指标之间具有更高的结构复杂性与异质性。分析结果不仅说明了问卷与行为任务之间低相关性的结构基础，也为后续在项目层对条目进行筛选和结构提取提供了结构层面的支持和理论铺垫。





七、条目筛选与efa分析

为提升因子结构的清晰度与条目的结构贡献度，这部分研究首先采用条目共性与随机基线比较法对原始条目进行筛选。具体地，通过计算每个条目在真实数据下的共性，并与相同结构的随机数据中该条目的共性分布进行比较，仅保留共性高于随机分布95%分位的条目（即在结构贡献上显著强于随机性的条目）。此方法有效排除了结构不明确或解释力较弱的变量，最终保留了 140 个结构贡献显著的条目。随后，使用平行分析、MAP检验与BIC等方法综合判断因子数，结果显示建议因子数在8～11之间。综合结构简洁性与模型拟合指标，最终提取8个因子，采用最小残差法（minres）与oblimin斜交旋转进行EFA分析。结果显示大多数条目在某一因子上具有清晰主加载，结构聚合良好，仅少数条目因载荷不足被进一步剔除。最后，通过因子–条目聚类热图直观展示了条目在因子空间中的结构分布，验证了条目归属的合理性与因子结构的可解释性。



结合前述相关分析与因子结构提取的结果可以看出，积极自我认知可能并非单一维度构念。在不同测量方式之间相关性较低的同时，结构性分析进一步表明，在共性筛选后的条目中存在多个聚合良好的因子维度。该结果提示，个体的积极自我可能由多个侧面构成，具有内在结构复杂性。

1. 预测分析

为评估积极自我主因子对多项心理健康指标的预测效度，这部分研究构建了线性与非线性回归模型，以因子得分预测四项心理变量（SGPS、PHQ、GAD、SWB）。模型训练与评估过程如下：

1. 数据准备

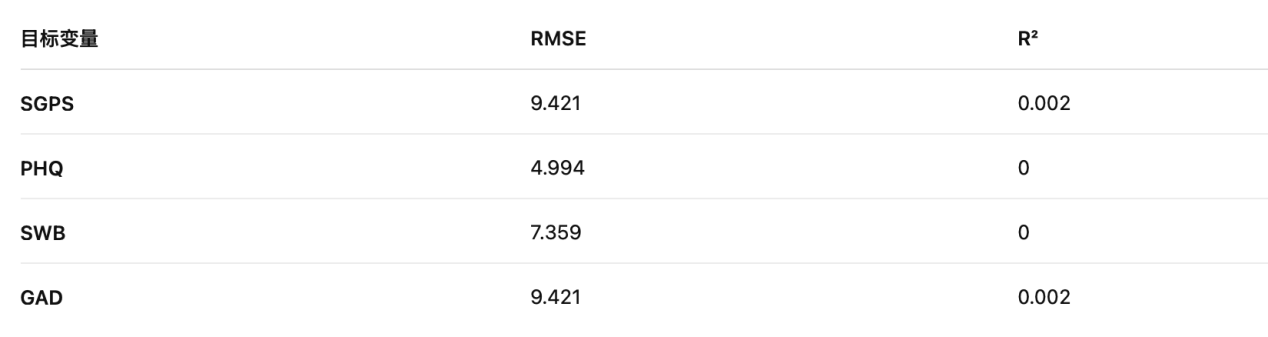
（1）预测变量：基于 EFA 提取的前4个主成分因子得分（MR1–MR4），作为特征输入；（2）目标变量：四项心理健康指标，包括 SGPS（积极自我总分）、PHQ（抑郁）、GAD（焦虑）和 SWB（主观幸福感）；

（3）数据分割：将原始样本按照 7:3 的比例划分为训练集与测试集，保证测试样本独立性。

2.模型选择

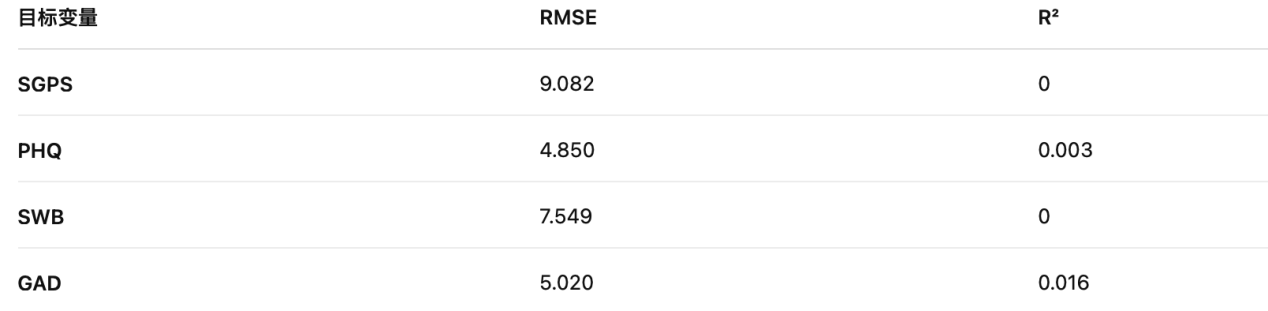
（1）线性模型

我们使用 Lasso 回归模型，以积极自我主成分因子（EFA 提取的前4个因子得分）为预测变量，分别预测四个心理健康变量（SGPS、PHQ、GAD、SWB），在 30% 测试集中评估预测效果。结果如下：所有变量的 R² 均接近于 0，说明基于主成分因子的 Lasso 模型无法有效解释心理指标的方差。这表明主成分因子与心理健康变量之间不存在稳定的线性关系，我们需要进一步探索中介变量或非线性心理机制（如元认知、自我调节信心）才能揭示更深层的关系。



1. 非线性模型

为了探索潜在的非线性关系，我们进一步采用随机森林模型进行建模，并使用相同的因子得分预测 SGPS、PHQ、GAD 与 SWB。结果如下：与 Lasso 模型一致，随机森林模型也未显示出明显的预测优势，R² 值同样非常接近 0，说明即便引入复杂的非线性结构和交互关系，也无法提升模型对心理健康变量的解释力。



尽管我们从积极自我条目中提取了结构清晰的因子用于建模，无论采用线性方法（Lasso）还是非线性方法（随机森林），都无法有效预测个体的心理健康水平。所有模型的 R² 值均趋近于 0，说明积极自我结构层面上的得分可能并不直接反映个体当下的心理适应状态。这一结果强调了自我结构的认知性描述与功能性预测之间存在明显脱节。本次分析中，所有主成分变量在四个目标变量模型中的 %IncMSE 均整体偏低，大多集中于 0–5% 区间，部分甚至为负值，说明这些因子整体上在预测任务中贡献较小，模型对其依赖程度有限。然而，值得注意的是，在不同预测模型中，不同主因子的 %IncMSE 表现出了一定的方向性和特异性。



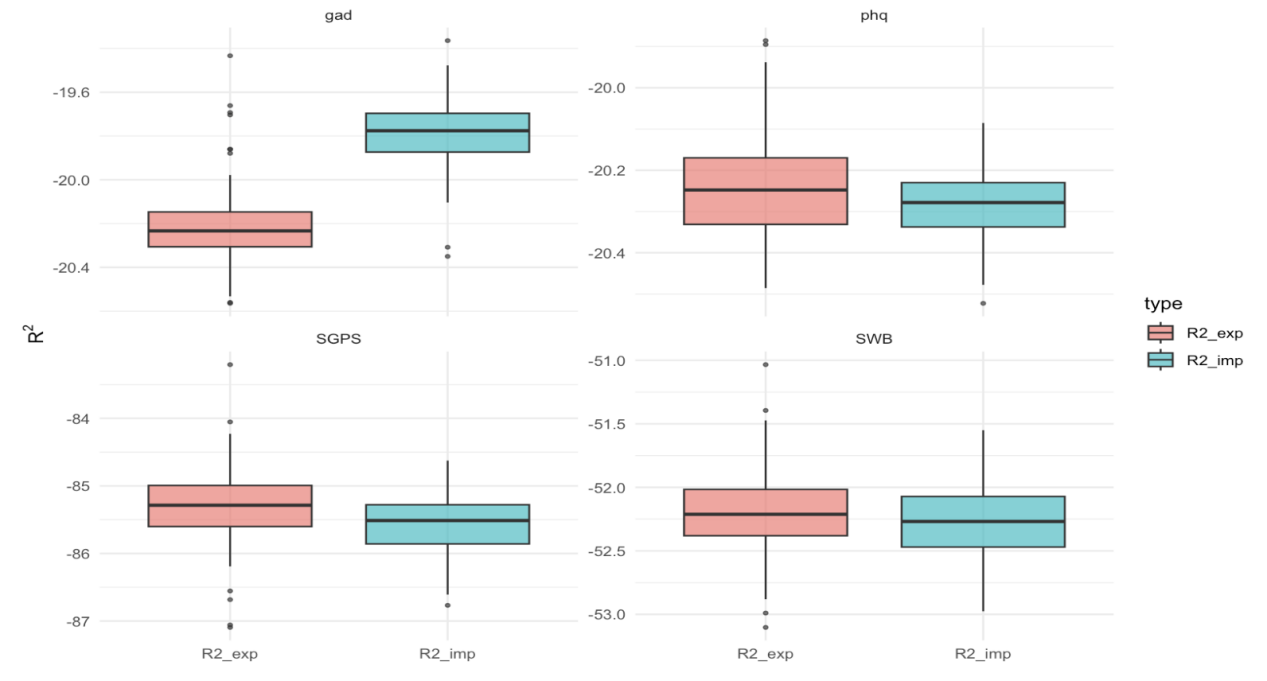
除了以上分析外，我们还采用了bifactor分析，在项目层提取共同因子和特异因子，但模型拟合效果较差；采用bimodel在项目层试图提取行为任务与自我报告两个主成分，模型也并未拟合；使用贝叶斯因子方法筛选条目，在调整参数和链数后模型未达到收敛；采用共性条目筛选后，行为任务均被剔除，使用在上述的分析中采取了其他方法。

九、“结构—功能悖论”分析

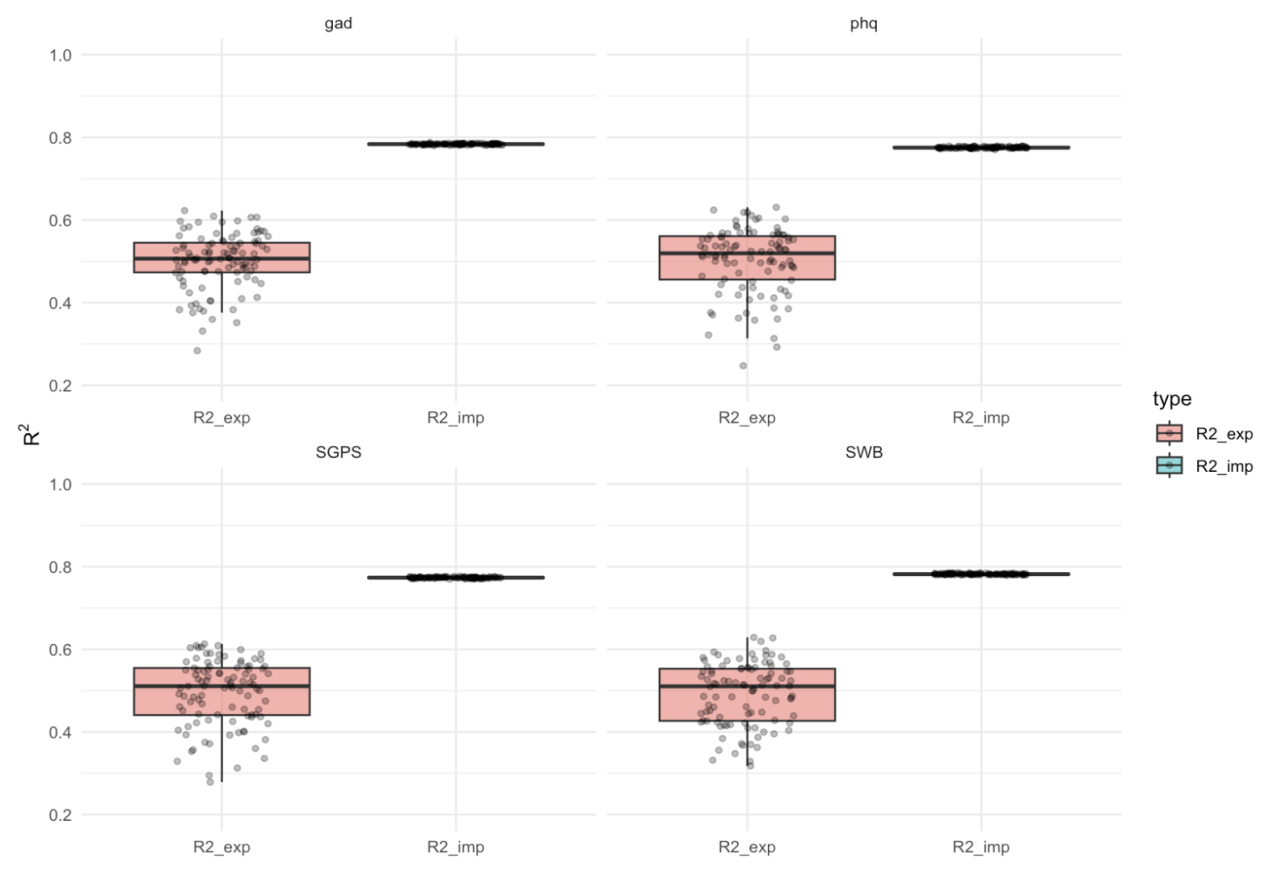
为比较显性测量（自我报告问卷）与隐性测量（行为任务）在心理健康变量预测中的表现，本研究分别基于线性模型（LASSO回归）与非线性模型（随机森林）开展了系统建模分析。但当预测心理健康指标时，显性测量和隐性测量数量严重不平衡，如何做出公平、可解释的对比，从而排除显性测量在“数量”上占据的优势成为数据分析面临的一大难题。

因此在分析前期的条目选择上，我们采取了“等量抽样 + 多次重复建模”的方法。具体思路如下：1.从133个显性条目中随机抽取7项；2.与7项隐性变量分别建模；3.重复此过程（100次）；4.对比平均R²、RMSE、变量重要性分布等。

在LASSO模型中，我们发现，无论是显性测量还是隐性测量，所有 R² 都是负数，这说明LASSO线性模型几乎没有预测力，两种测量方式在功能性预测力上表现都非常有限，甚至不足以超过零基线，显性测量在大多数变量上略强于隐性测量，但优势不明显，同时说明线性模型难以充分利用行为任务所包含的潜在预测信号。



而在随机森林模型中，预测效度显著提升，尤其是隐性测量在所有心理健康变量上表现出高度一致的预测力，R²稳定维持在0.78左右，表现出远高于显性测量的准确性与稳定性。显性测量在非线性建模下的预测力亦有所提升（R²≈0.5–0.6），但仍呈现出一定的抽样波动。整体而言，行为任务作为隐性测量，尽管信效度检验时表现较差，结构上也并未表现出清晰的结构，却在功能性预测中表现出更高的效度与更强的稳定性。



这些结果共同支持了“结构—功能悖论”的理论观点，即结构化程度较高的显性测量并不必然具备更强的功能性预测力，而结构松散的隐性测量在复杂建模框架中可能展现出更优的实用价值。