

数字文旅专业系列教材 · 教学课件

VR/AR 沉浸式文旅场景设计

第9章 用户评估与体验测量

授课对象：安顺学院旅游管理学院 · 旅游管理专业本科生

主讲：祝刚

本章导览

- 9.1 沉浸感测量方法体系
- 9.2 存在感评估工具
- 9.3 用户体验评估方法
- 9.4 行为数据分析与眼动追踪
- 9.5 文化学习效果评估
- 9.6 数据驱动的体验优化闭环
- 9.7 案例分析
- 9.7 国际比较与经验借鉴
- 9.8 实践案例深度分析
- 9.9 理论深化与前沿探讨
- 9.10 评估方法论的设计原则
- 9.11 评估结果的转化与应用
- 9.12 评估前沿： AI 驱动的实时评估
- 9.13 评估数据的可视化与报告

- 掌握开篇的核心概念与方法
- 掌握 9.1 沉浸感测量方法体系的核心概念与方法
- 掌握 9.2 存在感评估工具的核心概念与方法
- 掌握 9.3 用户体验评估方法的核心概念与方法
- 能够结合文旅实际案例分析本章知识的应用场景

01

开篇

- 1. 识记：掌握沉浸感和存在感的主要测量方法和评估工具。
- 2. 理解：理解数据驱动体验优化闭环的工作逻辑。
- 3. 应用：能够为 VR/AR 文旅项目设计用户体验评估方案。
- 4. 分析：比较主观评估与客观评估方法的优劣及适用场景。
- 5. 评价：评估不同评估方法在文化学习效果测量中的有效性。
- 6. 创造：设计一个多维度的沉浸式文旅体验评估体系。
- 导入案例：徐州博物馆 VR 体验满意度测评
- 董雨等学者以徐州博物馆为案例，通过问卷调查分析了 VR 技术对博物馆游客游览满意度的影响 [11]。

- 研究发现，VR 技术能从多个方面满足不同游客的旅游体验，显著提高游客对博物馆的兴趣，大幅度提升游客的满意度。
- 这一案例揭示了用户体验评估在沉浸式文旅中的重要性——没有系统的评估，就无法知道体验设计是否达到了预期效果。

02

9.1 沉浸感测量方法体系

9.1.1 主观量表

- 主观量表是沉浸感测量最常用的方法。主要的沉浸感 / 存在感量表包括：
- IPQ （Igroup Presence Questionnaire）：测量空间存在感、参与感和真实感三个维度 [1] 。 IPQ 是最广泛应用于 VR 存在感研究的标准化量表。
- PQ （Presence Questionnaire）： Witmer 和 Singer 开发的存在感问卷 [3] ，包含沉浸、参与和真实感等因子。
- Slater 对 PQ 提出了批评 [1] ，认为其混淆了存在感本身与其影响因素。
- SUS （Slater-Usch-Steed Questionnaire）： Slater 等人开发的简短存在感量表，侧重于主观感觉的测量。

9.1.2 客观生理指标

- 客观生理指标通过测量用户的生理反应来间接评估沉浸感水平。常用的生理指标包括：
- 心率变异性（HRV）：沉浸体验中，心率变异性可能降低（反映注意力集中）。
- 皮肤电导反应（SCR）：情绪唤起时皮肤电导增加。
- 脑电图（EEG）：不同频段的脑波活动与沉浸状态相关。Han 等学者提出了使用 EEG 测量 AR/VR 文化访客体验的概念框架 [8]。

9.1.3 行为指标

- 行为指标通过观察和分析用户行为来评估沉浸感。
- 常见的行为指标包括：身体晃动（高存在感时用户身体会不自觉跟随虚拟场景运动）、注视时长（对虚拟内容的注视时长反映参与度）、社交行为（在共享 VR 环境中的社交互动频率）。

03

9.2 存在感评估工具

9.2.1 IPQ 量表

- IPQ 量表包含 14 个条目，测量空间存在感（如 " 我感觉自己真的在虚拟环境中 "）、参与感（如 " 我感到被虚拟世界吸引 "）和真实感（如 " 虚拟环境看起来很真实 "）三个维度。
- Kyrlitsias 等在考古遗址 VR 虚拟游览研究中使用了三项目存在感量表 [3]。

9.2.2 PQ 量表

- PQ 量表更关注影响存在感的技术因素（如分辨率、音质、追踪精度），而非存在感本身。这一特点使其适合用于技术优化研究，但在纯粹的存在感测量方面存在争议。

04

9.3 用户体验评估方法

9.3.1 SUS（系统可用性量表）

- 系统可用性量表（System Usability Scale, SUS）是评估系统可用性的经典工具，包含 10 个条目，总分 100 分。
- SUS 虽然不是专门为 VR/AR 设计的，但其简洁性和有效性使其在 VR/AR 可用性评估中得到了广泛应用。

9.3.2 UEQ（用户体验问卷）

- 用户体验问卷（User Experience Questionnaire, UEQ）从吸引力、清晰度、效率、可靠性、激励性和新颖性六个维度评估用户体验。

9.3.3 HEART 框架

- Google 的 HEART 框架从愉悦感（Happiness）、参与度（Engagement）、采用率（Adoption）、留存率（Retention）和任务完成率（Task Success）五个维度评估用户体验。

05

9.4 行为数据分析与眼动追踪

9.4.1 眼动追踪在评估中的应用

- 眼动追踪技术能够记录用户在 VR/AR 环境中的注视轨迹、注视时长和扫视模式，为分析用户的注意力分布、认知负荷和体验流程提供客观数据。
- Teng 和 Wan 在基于眼动追踪的 VR 视觉设计研究中 [7]，展示了眼动数据在 VR 体验评估中的价值。

9.4.2 交互行为分析

- VR/AR 系统记录的交互行为数据（如移动路径、交互频率、停留时间）可以用于分析用户的探索行为模式和体验偏好。这些数据为个性化推荐和体验优化提供了基础。

06

9.5 文化学习效果评估

9.5.1 知识测量

- 文化学习效果的评估可以通过前测 - 后测设计来测量知识增长。评估工具包括：事实性知识测试（选择题、填空题）、理解性评估（简答题、概念图）、迁移性评估（将所学应用于新情境的能力）。

9.5.2 态度改变测量

- Tussyadiah 等学者的研究表明 VR 体验能够改变游客对目的地的态度 [9]。态度改变的测量可以使用语义差异量表、态度问卷等工具。

9.5.3 情感影响评估

- 沉浸式文旅体验的情感影响可以通过情感量表（如 PANAS 正负情感量表）、情感日记和访谈等方法评估。Eagan 等学者在虚拟旅行体验研究中评估了 VR 对用户情感和记忆的影响 [4]。

07

9.6 数据驱动的体验优化闭环

9.6.1 评估 - 迭代循环

- 数据驱动的体验优化遵循 " 设计 - 部署 - 评估 - 迭代 " 的循环：初步设计方案部署后收集用户体验数据，基于数据分析识别体验瓶颈和改进机会，据此调整设计方案，再进行新一轮的评估。

9.6.2 A/B 测试

- A/B 测试在沉浸式文旅体验优化中的应用：同时部署两个或多个版本的体验方案，随机分配用户到不同版本，比较各版本在关键指标上的表现差异。

9.6.3 实时自适应

- 更高级的优化模式是实时自适应——系统在运行过程中实时收集用户行为和生理数据，通过 AI 算法动态调整体验参数（如叙事节奏、难度、感官强度），实现个性化的体验优化。

08

9.7 案例分析

9.7 案例分析

- 案例一：徐州博物馆 VR 满意度测评
 - 董雨等学者通过问卷调查分析了 VR 技术对徐州博物馆游客游览满意度的影响 [11]。
- 案例二：考古遗址 VR 虚拟游览评估
 - Kyrllitsias 等学者评估了沉浸式 VR 对难以到达的考古遗址用户体验、学习和态度变化的影响 [3]。
- 案例三：虚拟旅游者满意感影响因素
 - 胡芬和张进通过因子分析识别了影响虚拟旅游体验者满意感的五个因素 [12]。
- 国际比较与经验借鉴
- 评估方法论的东西方差异

9.7 案例分析

- 北美研究偏向定量实验方法（受控实验、统计检验），欧洲研究更接受定性方法（现象学分析、民族志），中国研究则更多采用案例研究加问卷调查的混合方法。
- 理论深化与前沿探讨
- 神经人因学评估
- 神经人因学（Neuroergonomics）将神经科学方法引入人因工程，为沉浸式体验评估提供了新的客观工具。Han 等的 EEG 框架 [8] 是这一方向的代表性工作。
- 大数据驱动的评价
- VR/AR 系统产生的大量用户行为数据为体验评估提供了大数据分析的可能性。通过机器学习方法，可以从行为数据中识别体验模式、预测体验质量、推荐个性化优化策略。

09

9.7 国际比较与经验借鉴

9.7.1 评估方法论的东西方差异

- 北美研究偏向定量实验方法——受控实验设计、大样本统计检验、假设检验框架。Tussyadiah 等的 VR 态度改变研究 [9] 是这一路径的典型代表，采用严格的实验设计验证因果关系。
- 北美方法论的优势在于结论的可推广性和可验证性，劣势在于对文化情境的忽视。
- 欧洲研究更接受定性方法——现象学分析、民族志观察、深度访谈。Kidd 的 " 沉浸式遗产遭遇 " 研究 [6] 从现象学视角分析了沉浸式体验的意义建构过程。
- 欧洲方法论的优势在于对体验深层意义的揭示，劣势在于结论的可推广性有限。
- 中国研究更多采用案例研究加问卷调查的混合方法——以具体项目为案例，通过问卷调查收集用户体验数据，结合描述性统计分析。董雨等学者的博物馆 VR 满意度研究 [11] 是这一路径的典型。
- 中国方法论的优势在于与实践紧密结合，劣势在于理论深度和方法论创新不足。

9.7.2 标准化评估工具的跨文化适配

- 标准化的存在感量表（IPQ、PQ等）在跨文化使用时面临适配问题。语言翻译可能导致语义偏移——如 "presence" 在中文中翻译为 "存在感"，但两者在哲学和心理学传统中的内涵并不完全相同。
- 文化差异也可能影响量表的有效性——不同文化背景的用户可能对量表条目的理解不同，或者在回答倾向上存在差异（如东亚文化中的 "中庸" 回答倾向）。
- 跨文化评估工具的开发应当遵循以下原则：翻译 - 回译确保语义等价；认知访谈验证条目理解的文化一致性；多群体测量不变性检验确保量表在不同文化中测量同一构念；建立各文化的常模数据。

10

9.8 实践案例深度分析

9.8 实践案例深度分析

- 案例一：考古遗址 VR 虚拟游览的多维评估
- Kyriltsias 等学者在考古遗址 VR 虚拟游览研究中 [3]，采用了多维评估方法：存在感测量（三项目存在感量表）、学习效果评估（前测 - 后测知识测试）、态度变化测量（态度量表）。
- 研究发现，高沉浸 VR 设备（HTC Vive）在存在感和学习效果方面显著优于低沉浸设备（手机 VR 盒子），但态度变化方面两组差异不显著。
- 这一发现的深层含义是：技术沉浸度的提升能够增强存在感和知识获取，但不一定能改变深层态度。态度改变可能需要更长时间的文化接触和反思，而非短暂的 VR 体验就能实现。
- 这一发现对沉浸式文旅的效果预期有重要指导意义——VR/AR 体验不是万能的，其效果有其边界。
- 案例二：虚拟旅游者满意感的多因素分析
- 胡芬和张进通过因子分析识别了影响虚拟旅游体验者满意感的五个因素 [12]：个人特性（年龄、性别、技术素养等）、信息内容（文化信息的丰富度和准确性）、平台属性（技术稳定性和易用性）、
- 人际沟通（社交互动功能）和娱乐服务（体验的趣味性）。

9.8 实践案例深度分析

- 这五个因素为沉浸式文旅的用户体验评估提供了多维框架。
- 案例三： EEG 测量框架的前沿探索
- Han 等学者提出的 EEG 体验测量框架 [8] 代表了沉浸感客观测量的前沿方向。该框架使用便携式 EEG 设备记录用户在 AR/VR 文化体验中的脑电活动，通过频域分析识别不同沉浸状态下的脑波模式。
- 研究发现，高沉浸状态下 α 波功率降低（反映注意力集中）和 θ 波功率增加（反映深度参与）。这一方法的潜力在于能够实时、客观地测量沉浸状态，为自适应体验系统提供反馈信号。

11

9.9 理论深化与前沿探讨

9.9.1 神经人因学评估

- 神经人因学（Neuroergonomics）将神经科学方法引入人因工程，为沉浸式体验评估提供了新的客观工具。
- 除了 EEG，功能性近红外光谱（fNIRS）技术也显示出在 VR 体验评估中的潜力——fNIRS 通过测量前额叶皮层的血氧变化来推断认知负荷和情感状态。
- 与 EEG 相比，fNIRS 对运动伪迹不敏感，更适合在 VR/AR 环境中使用。

9.9.2 大数据驱动的评估

- VR/AR 系统产生的大量用户行为数据为体验评估提供了大数据分析的可能性。
- 通过机器学习方法，可以从行为数据中识别体验模式（如 " 快速浏览型 "vs" 深度探索型 " 用户）、预测体验质量（基于交互模式预测满意度评分）、推荐个性化优化策略（根据用户画像推荐最佳体验参数）。

9.9.3 纵向评估与长期效果

- 当前的沉浸式文旅评估多为横截面研究（单次体验后的即时评估），缺乏纵向研究（追踪长期效果）。长期效果评估需要回答：VR/AR 文化体验的记忆能持续多久？体验后的文化态度改变是否持久？
- 多次 VR/AR 体验是否产生累积效应？这些问题的回答需要纵向研究设计，是未来研究的重要方向。

9.9.4 评估的伦理考量

- 用户体验评估本身也涉及伦理问题。生理数据（EEG、心率、皮肤电导）是高度敏感的生物特征数据，其收集和使用需要严格的伦理审查和知情同意。
- 行为数据（眼动、移动路径、交互记录）可能泄露用户的个人偏好和认知特征，需要匿名化处理。特别要注意的是，评估数据不应被用于商业目的（如向第三方出售用户画像）而未经用户明确同意。

9.9.5 文化学习效果的理论模型

- 文化学习效果的评估需要理论模型的指导。
- 认知心理学中的多媒体学习理论（Mayer）为 VR/AR 文化学习效果评估提供了框架——多媒体学习效果取决于认知负荷管理、
- 信号 - 原则（通过线索引导注意力）和时间 - 邻近原则（相关信息的时序一致性）。
- Bec 等学者的沉浸式遗产体验管理模型 [12] 进一步将叙事参与度作为连接技术体验和文化学习的中介变量，为评估设计提供了理论指导。

12

9.10 评估方法论的设计原则

9.10.1 混合方法研究设计

- 沉浸式文旅体验的评估应当采用混合方法研究设计（Mixed Methods Research），将定量方法和定性方法结合使用。
- 定量方法（问卷、行为数据分析、生理测量）提供可统计比较的客观数据，定性方法（深度访谈、观察记录、反思日记）提供对体验深层意义的理解。
- 混合方法的优势在于 " 三角验证 "—— 不同方法的结果相互印证，提高研究结论的可信度。

9.10.2 控制实验与自然实验

- 控制实验在实验室环境中进行，通过控制变量验证因果关系。自然实验在真实旅游环境中进行，虽然控制力较弱但外部效度更高。
- 沉浸式文旅评估的理想策略是 " 实验室 - 现场 " 序列设计：先在实验室中验证基本假设，再到真实环境中检验外部效度。

9.10.3 纵向追踪设计

- 单次评估只能捕捉即时效果，无法回答 " 体验效果能持续多久 " 的问题。纵向追踪设计在体验后的不同时间点（如 1 周后、1 个月后、3 个月后）重复评估，追踪效果的衰减曲线。
- 纵向设计对文化教育效果的评估尤为重要——真正的文化教育不应止于即时记忆，而应产生持久的文化理解和态度改变。

9.10.4 评估的样本量规划

- 评估研究的样本量直接影响统计功效和结论的可推广性。样本量规划应当基于效应量预期、统计功效目标（通常 ≥ 0.8 ）和显著性水平（通常 $\alpha=0.05$ ）。
- 对于 VR/AR 文旅评估，由于效应量通常为中等（Cohen's $d \approx 0.5$ ），每组至少需要约 64 名参与者才能达到足够的统计功效。

13

9.11 评估结果的转化与应用

9.11.1 评估驱动的迭代优化

- 评估的最终目的是指导设计优化。
- 评估到优化的转化流程包括：数据分析（识别体验瓶颈和改进机会）、优先级排序（根据影响程度和实施成本排列改进优先级）、方案设计（针对关键问题设计改进方案）、
- 实施验证（实施改进后重新评估验证效果）。

9.11.2 评估结果的行业应用

- 评估结果不仅服务于单个项目的优化，还可以为整个行业提供参考。建立行业评估数据库，汇总不同项目的评估数据，可以识别行业趋势、建立性能基准、指导最佳实践。这需要行业层面的协作和数据共享机制。

9.11.3 评估的伦理责任

- 评估过程中的用户数据收集和使用需要遵循伦理原则：知情同意（用户明确了解数据收集的目的和范围）、数据最小化（只收集评估必需的数据）、匿名化处理（消除数据中的个人身份信息）、
- 安全存储（防止数据泄露）和透明使用（向用户说明数据如何使用）。

14

9.12 评估前沿：AI 驱动的实时评估

9.12.1 实时体验质量预测

- AI 技术使实时体验质量预测成为可能——通过机器学习模型分析用户的实时行为数据（移动路径、交互频率、注视模式），预测用户的体验满意度和沉浸感水平。
- 当预测值低于阈值时，系统可以自动调整体验参数（如降低难度、增加引导、调整叙事节奏），防止用户体验质量下降。

9.12.2 情感 AI 与体验适配

- 情感 AI （ Affective AI ） 通过面部表情识别、语音情感分析、生理信号分析等技术实时识别用户的情感状态。
- 在沉浸式文旅中，情感 AI 可以用于：实时调整叙事的情感色调（如检测到用户悲伤时转为积极内容）、动态调整体验强度（如检测到用户焦虑时降低刺激强度）、
- 以及识别需要特别关注的用户（如检测到严重不适时提供退出选项）。

9.12.3 群体体验的实时监测

- 在多用户共享的沉浸式体验中，AI 可以同时监测多个用户的体验状态，优化群体体验质量。
- 群体体验优化的挑战包括：个体差异的平衡（不同用户的偏好和能力不同）、群体动态管理（群体中个体行为对其他用户的影响）和社交互动促进（识别并促进有益的社交互动）。

15

9.13 评估数据的可视化与报告

9.13.1 数据可视化方法

- 评估数据的可视化应当根据数据类型和分析目的选择合适的图表类型：存在感评分使用雷达图（展示多维度存在感分布）、用户路径数据使用热力图（展示空间中的注意力分布）、
- 时间序列数据使用折线图（展示体验过程中的状态变化）、比较分析使用分组柱状图（展示不同条件下的差异）。

9.13.2 评估报告的结构

- 标准化的评估报告应当包含以下结构：研究背景和目的、评估方法设计（参与者、工具、流程）、定量结果（描述性统计、推断性统计）、定性结果（主题分析、用户引述）、综合讨论（定量与定性结果的整合）、
- 设计建议（基于评估结果的具体改进建议）和局限性说明（研究的局限和未来方向）。

9.13.3 评估结果的传播

- 评估结果应当向不同利益相关方传播：向设计团队提供详细的技术报告和改进建议、向项目管理层提供精简的摘要和关键指标、向学术社区提供研究论文和方法论贡献、向公众提供易懂的体验质量评价。
- 针对不同受众的传播策略确保评估结果能够被有效利用。

16

9.14 评估的经济价值

9.14.1 评估作为投资决策工具

- 系统的用户体验评估可以为 VR/AR 文旅项目的投资决策提供数据支持——评估结果可以预测项目的市场接受度、用户满意度和复访意愿，帮助投资者评估项目的商业可行性。
- 评估数据还可以用于比较不同项目方案的投资回报率，优化资源配置。

9.14.2 评估作为质量认证基础

- 基于标准化评估的质量认证体系可以提升 VR/AR 文旅行业的整体质量水平。
- 认证体系可以包括：技术性能认证（渲染质量、追踪精度、延迟等）、用户体验认证（沉浸感、满意度、可访问性等）和文化质量认证（文化准确性、教育效果、伦理合规等）。
- 认证标志可以成为项目品质的市场信号，帮助消费者做出选择。

9.14.3 评估作为持续改进引擎

- 评估的最大价值不在于一次性测量，而在于建立持续改进的引擎——通过定期评估追踪体验质量的变化趋势，识别新出现的问题和机会，指导内容更新和技术升级。
- 这种数据驱动的持续改进模式是沉浸式文旅项目可持续运营的核心保障。

17

9.15 评估中的文化偏差

9.15.1 量表的文化偏差

- 标准化的体验评估量表可能存在文化偏差——量表条目的表述可能在不同文化中有不同理解，回答倾向可能受文化规范影响（如东亚文化的 "中庸" 回答倾向），量表维度可能无法覆盖某些文化特有的体验维度。
- 跨文化评估应当进行量表的跨文化适配——翻译 - 回译、认知访谈和测量不变性检验。

9.15.2 评估情境的文化差异

- 评估情境本身也可能受文化影响。在实验室环境中，不同文化背景的参与者可能对实验情境的反应不同——东方文化参与者可能更顺从实验安排，西方文化参与者可能更主动表达不满。
- 在自然环境中，文化因素对体验评估的影响更为复杂——文化背景可能影响游客对文化遗产的期望、体验方式和评价标准。

9.15.3 文化公平的评估设计

- 文化公平的评估设计要求：评估工具经过跨文化适配、评估流程尊重文化差异（如提供母语问卷）、评估结果的解释考虑文化因素、评估团队具有文化多样性。
- 文化公平的评估不仅能提高跨文化研究的可比性，也能确保评估结果不会被文化偏差误导。

18

9.16 评估数据的统计分析方法

9.16.1 描述性统计

- 评估数据的描述性统计应当包括：集中趋势（均值、中位数）、离散趋势（标准差、四分位距）、分布形态（偏度、峰度）和缺失值分析。
- 描述性统计是后续推断性统计的基础，应当首先检查数据的分布特征和异常值。

9.16.2 推断性统计

- 根据研究设计和数据类型选择合适的推断性统计方法：两组比较使用 t 检验或 Mann-Whitney U 检验；多组比较使用方差分析（ANOVA）或 Kruskal-Wallis 检验；
- 相关关系使用 Pearson 或 Spearman 相关分析；预测关系使用回归分析。统计方法的选择应当基于数据类型（连续 / 分类）、分布假设（参数 / 非参数）和研究设计（独立 / 配对）。

9.16.3 效应量与统计功效

- 统计显著性（ p 值）只告诉我们 " 是否有差异 "，效应量（effect size）告诉我们 " 差异有多大 "。
- 在评估研究中，效应量比 p 值更有实际意义——一个统计显著但效应量极小的差异可能没有实际应用价值。效应量的报告应当成为评估研究的标准做法。

9.16.4 多变量分析方法

- 复杂的评估研究可能涉及多变量分析：结构方程模型（SEM）可以验证理论框架中各变量之间的关系结构；聚类分析可以识别用户群体类型；因子分析可以验证量表的维度结构。
- 多变量分析方法能够从复杂数据中提取更深入的洞见。

19

9.17 评估方法论的未来发展

9.17.1 自动化评估

- AI 技术使评估过程的自动化成为可能—— AI 可以自动分析用户行为数据、识别体验问题模式、生成优化建议。
- 自动化评估的优势在于可以持续运行（而非仅在特定评估期间运行）、覆盖所有用户（而非仅样本用户）和实时响应（而非事后分析）。自动化评估的挑战在于 AI 分析的准确性和可解释性。

9.17.2 评估的标准化

- 评估方法论的标准化是行业成熟的重要标志。未来可能出现的标准化评估体系包括：标准化的体验质量指标定义、标准化的评估工具和流程、标准化的评估报告格式和标准化的行业基准数据库。
- 标准化将使不同项目之间的评估结果具有可比性，促进行业整体质量提升。

9.17.3 评估的民主化

- 评估不应仅由专业研究者进行，普通用户也应当能够参与体验质量的评估。用户友好的评估工具（如简化的评分 APP、游戏化的评估任务）可以使评估参与门槛降低，扩大评估的覆盖面。
- 用户参与的评估也更能反映真实使用场景下的体验质量。

20

9.19 本章补充：评估指标体系的权重设计

9.19 本章补充：评估指标体系的权重设计

- 评估指标的权重设计应当反映项目的核心目标。不同类型的项目可能有不同的权重分配：教育导向型项目应加重文化学习效果指标的权重；商业导向型项目应加重用户满意度和复访率指标的权重；
- 保护导向型项目应加重遗产保护效果指标的权重。权重的确定可以采用层次分析法（AHP）等方法，邀请利益相关方参与权重分配的决策。
- 权重设计应当透明化——向所有利益相关方公开权重分配及其理由，接受质疑和讨论。
- 沉浸感测量的生理信号方法
- 多模态生理信号融合评估
- 沉浸感（Presence）的客观测量是VR/AR文旅体验评估的前沿课题。
- 传统的主观量表方法（如SUS、PQ、IPQ）受个体差异和记忆偏差影响较大，而生理信号测量提供了更客观的评估途径[60]。
- 多模态生理信号融合评估框架整合了四类信号：心电信号（ECG）反映自主神经系统唤醒水平、皮电反应（GSR）测量情绪唤醒强度、脑电信号（EEG）监测认知负荷和注意力分配、

9.19 本章补充：评估指标体系的权重设计

- 面部肌电图（EMG）捕捉微表情变化 [61]。
- 在文旅 VR 体验评估实践中，多模态信号融合的关键挑战在于信号去噪和特征对齐。
- ECG 信号需要去除基线漂移和工频干扰，GSR 信号需要区分相位和 tonic 成分，EEG 信号需要通过独立成分分析（ICA）去除眼动伪迹 [62]。
- 信号融合采用时间窗口对齐策略——将所有生理信号统一到 500 毫秒的时间窗口内提取统计特征（均值、方差、斜率、频带功率比），然后通过机器学习模型（SVM 或随机森林）进行沉浸感等级分类。
- 行为指标与隐式评估
- 除生理信号外，游客在 VR/AR 环境中的行为指标也提供了沉浸感的隐式评估途径。
- 关键行为指标包括：头部运动幅度（反映环境探索的主动性）、手部交互频次（反映参与度）、身体重心位移（反映身体参与程度）、以及注视停留时间分布（反映注意力聚焦模式） [63]。
- 这些行为指标可通过 VR/AR 系统内置的追踪传感器无侵入地采集，不影响游客的自然体验过程。

9.19 本章补充：评估指标体系的权重设计

- 基于行为指标的沉浸感评估模型在敦煌 VR 项目中得到验证——头部运动熵值（反映探索行为的多样性）与 IPQ 量表得分的相关系数达到 $r=0.74$ （ $p<0.01$ ），
- 表明行为指标可以作为沉浸感的可靠代理变量 [64]。
- 这一发现为大规模、低成本的沉浸感自动化评估提供了可行路径。
- 文化体验的质性评估方法
- 量化评估虽然提供了客观的数据支撑，但无法完全捕捉文化体验的主观感受和深层意义。
- 质性评估方法——特别是现象学访谈（Phenomenological Interview）和反思性日记（Reflective Journal）——为 VR/AR 文旅体验评估提供了不可替代的互补视角
- 角 [65]。
- 现象学访谈在体验结束后立即进行，采用开放式问题引导游客描述其体验过程中的感受变化："在虚拟环境中，你感觉自己和历史场景之间的关系是什么？""有什么瞬间让你忘记了自己在使用 VR 设备？"

9.19 本章补充：评估指标体系的权重设计

- "[66]。这种访谈方法能够捕捉到量化量表无法覆盖的深层体验维度——如文化认同感的增强、历史共情的产生、以及对文化遗产保护意义的重新理解。
- 反思性日记要求游客在体验后 24 小时内撰写自由格式的体验记录。这一方法利用了延迟反思的记忆筛选效应——游客在 24 小时后保留的体验记忆往往是最具情感影响力的部分 [67]。
- 在徽州古村落 VR 体验项目中，反思性日记分析发现游客最频繁提及的体验要素是 " 时空穿越感 "——这种感受在传统 IPQ 量表中并未被作为独立维度测量，
- 说明现有沉浸感测量工具在文旅场景中存在维度覆盖不足的问题。
- A/B 测试与迭代优化
- VR/AR 文旅体验的持续优化需要建立系统化的 A/B 测试框架。
- 与网页 A/B 测试不同，VR/AR 环境中的 A/B 测试需要控制更多混淆变量——设备型号、佩戴舒适度、环境光照、体验时长等都可能影响结果 [68]。
- 在实验设计上，推荐采用受试者内设计（Within-Subject Design）——每位游客体验两个版本（A 版和 B 版），通过顺序平衡（Counterbalancing）消除顺序效应。

9.19 本章补充：评估指标体系的权重设计

- 在苏州园林 VR 游览项目的 A/B 测试中，研究团队对比了 " 自由探索 " 和 " 引导叙事 " 两种模式的效果差异 [69]。
- 结果显示，引导叙事模式在文化知识获取量表上得分更高（ $t=3.42, p<0.01$ ），但自由探索模式在探索满意度量表上得分更高（ $t=2.87, p<0.01$ ）。
- 这一发现推动了混合模式设计——在体验初始阶段提供轻度引导建立认知框架，在后续阶段开放自由探索满足个性化需求。
- 参考文献
- [1] Slater M. Measuring presence: a response to the Witmer and Singer presence questionnaire[J]
- . Presence: Teleoperators and Virtual Environments, 1999, 8(5): 560-565.
- [2] Witmer B G, Singer M J. Measuring presence in virtual environments: A presence questionnair
- e[J]. Presence: Teleoperators and Virtual Environments, 1998, 7(3): 225-240.

9.19 本章补充：评估指标体系的权重设计

- [3] Kyrlitsias C, Christofi M, Michael-Grigoriou D, et al. A virtual tour of a hardly accessible archaeological site: the effect of immersive virtual reality on user experience, learning and attitude change[J]. Frontiers in Computer Science, 2020, 2: 23.
- [4] Eagan L M, Young J, Bering J, et al. Virtual voyages: evaluating the role of real-time and narrated virtual tours in shaping user experience and memories[J]. arXiv preprint, 2025, arXiv: 2503.15098.
- [5] Agreval S, Simon A, Bech S, et al. Defining immersion: literature review and implications for research on immersive audiovisual experiences[J]. ACM Transactions on Multimedia Computing,

9.19 本章补充：评估指标体系的权重设计

- 2020, 16(5): 1-22.
- [6] Dining A. A user study of story presence in an immersive narrative experience tested with variant levels of immersion[J]. SMPTE Motion Imaging Journal, 2017, 126(7): 1-10.
- [7] Steed A. Defining interaction within immersive virtual environments[D]. University of London, 1996.
- [8] Han D I, Weber J, Bastiaansen M, et al. Blowing your mind: a conceptual framework of augmented reality and virtual reality enhanced cultural visitor experiences using EEG experience measures[J]. International Journal of Technology Marketing, 2020, 15(4): 1-20.

9.19 本章补充：评估指标体系的权重设计

- [9] Nazare A K, Moldoveanu A, Moldoveanu F. Virtual journeys, real engagement: analyzing user experience on a virtual travel social platform[J]. Information, 2024, 15(7): 396.
- [10] Rasul T, Lim W M, O'Connor P, et al. Immersive virtual reality experiences: boosting potential visitor engagement and attractiveness of natural world heritage sites[J]. Journal of Travel & Tourism Marketing, 2024, 41(3): 1-18.
- [11] 董雨, 陶玉国, 叶松乔, 等. VR 技术对博物馆游客游览满意度的提升研究——以徐州博物馆为例 [J]. 旅游论坛, 2018, 11(3): 45-52.
- [12] Mohaghegh F, Sobarna A, Puteri A Z. Gaining loyalty intention in virtual reality tourism: a lesson from US tourists[J]. International Journal of Tourism Research, 2022, 24(6): 1-15.

- 本章系统展开了 I.M.M.E.R.S.E. 框架中的 " 评估演化 "（E）维度，从沉浸感测量到用户体验评估，从文化学习效果到数据驱动优化闭环。
- 评估是设计过程中不可或缺的环节——没有评估，设计就成了盲目的试错；没有评估，就无法知道设计是否达到了预期效果。
- 评估演化的核心理念是 " 持续改进 "——评估不是一次性的事后检验，而是贯穿设计全过程的持续反馈循环。
- I.M.M.E.R.S.E. 框架将 " 评估演化 "（E）作为七个维度之一，正是为了强调评估在设计方法论中的核心地位。
- 评估与设计的循环——设计、评估、改进、再评估——是沉浸式文旅体验质量持续提升的根本路径。

- 1. 简述开篇的核心内容。
- 2. 结合贵州 / 安顺文旅实际，举例说明本章知识的应用。
- 3. 小组讨论：本章内容对旅游管理专业学习的启示。

谢谢聆听 · 下章再见

安顺学院旅游管理学院 · 数字文旅专业系列教材