

物理教育游戏《粘粘世界》的学习性分析

The Analysis of Learnability——Illustrated by the Case of Educational Game *World of Goo*

王璇 Wang Xuan

内容摘要：目前，游戏经常被当作娱乐工具而忽视了它的趣味性学习体验，这种误区造成了游戏和学习的对立。本文以物理教育游戏为例，结合詹姆斯·保罗·吉的学习原则，使用回放分析引擎（RAE）回放游戏环境中的用户操作记录，使用获取与分析用户行为的工具包记录游戏过程中的指标，通过分析用户与教育目标的学习性与对齐性问题，及时发现游戏设计中的缺陷，以促进教育游戏发展。

关键词：教育游戏、学习性分析、对齐性分析、贝叶斯知识追踪

DOI:10.16272/j.cnki.cn11-1392/j.2016.04.015

教育游戏的优势体现在两方面：一是能够提高用户学习积极性；二是可以降低他们在真实环境中实践新技能所遭受的风险。这种优势使研究者开始建立一系列的教育游戏体系。^[1]斯洛塔·杨在测量中使用严肃方法证明游戏中教育的有效性。^[2]用户游戏过程体现其多变性的行为，测量方法也应随游戏不断迭代发展。在用户行为方面，每种方法都可能存在差异，研究者与开发者需运用多种方法测量。埃里克

等人开发了一种在教育游戏中获取与分析用户行为的工具包，可用于分析用户游戏过程中是否注意到学习知识点等问题，以此协助游戏设计师与研究人员开发能够教授儿童基本物理知识的游戏。^[3]

一、物理教育游戏

2003年，詹姆斯·保罗·吉（James Paul Gee）曾在《关于学习与文化，视频游戏教会我们什么》中指出，游戏作为一种高度可见媒体（Highly Visible Medium）与学校原本预想教授的知识、技能思想不同。物理教育游戏能够借助PC、移动终端创设情境、进行协作式学习和会话交流，成为用户主动学习、趣味探索的认知工具。如《粘粘世界》《蜡笔物理学》等游戏是通过模拟现实世界中的重力、对称性、重心、稳定性，在游戏过程中结合物理知识找出点过关。（图1）《粘粘世界》让用户使用液体分子构造桥梁、高塔等结构，同时要克服重力、峭壁、尖刺等不利因素，让液体分子通过搭建建筑吸入水管内。游戏中的液体分子始终不断移动，用户不仅需要保持建筑稳定，还要保障

目标数量的液体分子到达水管道（每搭建一步就会损失一个液体分子）。如果达到目标，用户就可以进入下一关，否则需要重新过关。

二、物理教育游戏的学习性测量方法

1. 物理教育游戏架构与初步学习性判断

《粘粘世界》的目标受众是8-15岁的青少年，以稳定性三原则为基础，即宽底、对称及低重心物体更稳定。通过在游戏过程中记录用户行为还原游戏过程，并分析其学习性，分析结果可用于迭代开发中改进游戏并完成学习性评估。运行环境在Unity引擎下完成，开发者可通过工具箱迅速修改变属性。英伟达物理运算引擎（Nvidia's PhysX engine）能够模拟现实世界中精确的物理环境。斯普林和佩莱格里诺将这类游戏定义为开放性游戏，开放性游戏由一系列规则支配游戏元素交互，使用户以非限定性结构解决问题，这使开放性游戏在教育层面上更具吸引力。^[4]詹姆斯·保罗·吉认为，游戏学习原则就是让用户承受较少实质性失败的情况下在仿真环境中大量实践。^[5]本文用该方法记录48位用户在《粘粘世界》中的行为及过关方式，通过初步比对发现：在重心更低更稳定为教育目标的关卡中，虽然大部分用户成功过关，但他们仍未抓住预设知识要点。图2左代表设计师预想的答案，而右边用户的搭建结果显然没有注意到该知识点。

2. 游戏中的学习性测量方法

教育游戏中的学习性由游戏环境外的转换量判断，通过控制用户游戏结果的外部测试法度量知识进展。但这种方法受时间与人力限制，需要在开发后期保存。理想状态是游戏内可评估的学习性测量，并辅以游戏外的学习性转化量对比研究，游戏内测量结果可为大数据收集增加更多途径。^[6]



1. 物理教育游戏

（1）ECD 分析法与贝叶斯知识追踪分析法

外文文献中，可搜索到的游戏内学习性测量法大部分是根据用户过关数量或记录完成的最后一关数据进行分析，这种结果直观且容易实现。戴利科鲁兹和贝克通过这种方法有效证明了游戏可以独立的测量用户数学能力。^[7]安德鲁森也用游戏中的单独一关测量了不同辅导形式学习的有效性。^[8]在回放《粘粘世界》用户游戏行为时，我们发现一些用户搭建的建筑在液体分子重力加速度情况下会倾斜到山坡一侧，倚靠着山的建筑水分子也可以到达水管道，这在游戏设置中并不算游戏失败。如果用户使用这种擦边球方法侥幸过关，便错过了重要的学习机会。为解决这个问题，游戏设计师曾尝试将山体移到离搭建区域更远的地方，使游戏反馈与教育目标能更好地结合。

还有一种较复杂的方法是以证据为中心的设计（Evidence Centered Design，以下简称 ECD）。ECD 由素质模型、证据模型和任务模型三部分组成：素质模型描述系统的目标知识；证据模型分析用户与搜索器定义题目获取得分系统的交互；任务模型搜集学习者系统内的表现。^[9]目前，舒特等人已将 ECD 成功运用到商业游戏中。贝叶斯知识追踪（Bayesian Knowledge Tracing，简称 BKT）也是教育游戏中比较流行的分析法。^[10]这种方式基于用户表现的知识原件，推断其学习知识状态，通过用户行为区别正确或错误的执行组件，并在贝叶斯网络中将行为排序后，决定用户理解每个知识组件的程度。这些方法需要不同的公式化数据、用户行为与晶粒度测试，当用户在游戏中表现得无经验时，分析结果与关卡表现结果相关。BKT 与 ECD 分析法的区别在于：BKT 需要分析单个用户的知识组件，ECD 可以进行多关卡分析。

测试使用用户行为数据中探寻与学习性测量相关的方法，发掘其他潜在方法。分析法包含两部分：记录用户行为日志的工具包和用户会话的回放分析引擎。目前，回放日志分析能将检测到的用户游戏会话作为游戏引擎中的状态即时实例化，这种分析法可避免重复运行庞大的研究数据。日志记录系统是匹兹堡科学学习中心（PSLC）数据仓库的 XML 格式，用于描述用户与智能辅导系统之间的交互。^[11]日志通过捕捉每关的用户行为，将有意义的行为



2.《粘粘世界》设计师预设方案与用户实际过关方案

定义为最小单元。这意味着将游戏世界置于情境化中，如放下或捡起一个物体，鼠标放在某（x,y）坐标轴上等。这些微小行为可以通过有效数据分析各种晶粒度，数据可粗略分析整个关卡表现，也可细致分析某一特殊关卡。在日志中，选择是用户行动的实体，被称为唯一标识符；行为是用户如何对实体行为进行文本描述的类型；输入是执行行动的参数。系统还可以记录用户间接行为。在《粘粘世界》中，当系统激活吸收水分子模式或计算剩余水分子时，日志便开始记录引擎行为。行为发生前，游戏描述状态与每个行为匹配，作为一个独立标识符实体与游戏行为配对。日志系统还使用行为前状态计算可能会移动实体的行为和具有忽略记录简单行为的能力，如按下“开始关卡”按钮就不予记录。《粘粘世界》游戏中，日志获取用户搭建的方向、速度、步数，以及通过的水分子数量，而两边悬崖的位置和下水管道参数是不变的，所以日志将其记录为初始状态。

有时，系统也未必能捕捉与分析法相关的行为，例如 ECD 分析法会认为行为与学习性测量无关，与游戏设置有关，这与游戏中发生的描述性日志行为动机是一致的。数据库由 C 语言编写，像开始或结束关卡这样的用户行为和特殊事件，实现器就会在函数中添加一条选择、行动和输入行为参数命令，它会保持与所有游戏实体相关的全局列表，实时更新并自动记录相关属性，还可以在设计时使用 Unity 引擎互动构建器，为实体添加脚本。这种高描述性日志一般都较大，只能使用智能分析，在《粘粘世界》中选择 48 位用户的两段游戏记录（从游戏开始到结束的数据），以每段 20 分钟、容量约 450MB 的体积记录到网络服务器中。但分析研究时这些数据却大到无法处理，因此只能使用局部记录。

（2）回放分析引擎

用户表现分析通过回放用户游戏会话完成，回放过程在第二个主要元件中实现，称为回放分析引擎（Replay Analysis Engine，以下简称 RAE）。RAE 也在 Unity 引擎中实现，读取用户游戏数据，并重建用户每一步行为会话。首先，RAE 构建行为状态，并让 Unity 引擎计算该行为结果，然后用 Prefab 预设架构复制或实例化复杂游戏对象，以确保分析师精准分析游戏对象，有效复制用户表现，引擎选项可决定何时停止模拟，进行分析。RAE 分析数据的方法有两种：第一种是应用程序接口（API）用数据库生成度量标准，进入用户行为创建的日志，方便分析师访问特定属性时使用。例如，《粘粘世界》中分析用户搭建的质心结构属性时，可用 API 检测实例化版本查找与下水管道连接的部分，计算出质心的平均数并创建整个结构的质心。第二种方法是用户日志预先分析，这一过程会产生统计分析表，使用 RAE 提供的 API 为日志过滤器开发通用测量法，比如输出特定的关卡或用户，引擎在运行表中记录数据时可以使用这些通用过滤器。RAE 将开发过程与测量方法、分析及迭代分离。

三、物理教育游戏《粘粘世界》中的学习性分析

《粘粘世界》的数据来自厦门市某学校 1-5 年级用户的课堂学习（年龄约 8-14 岁）。我们选择了两组游戏内前后测试关卡，选择标准为常规建筑关卡，这些关卡与正常关卡的区别是去掉了恢复前一步，也就是说用户错误操作无法恢复上一步操作。受试者在测试前要先玩教学关卡，以对比用户在获得大量游戏经验之前搭建的建筑与实践操作后搭建技能的变化。除参与游戏内评估外，受试者还要参加游戏外的纸质测试（游戏前和游戏后测试），包括底宽、

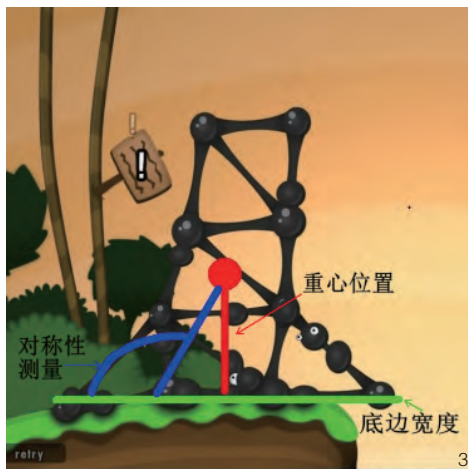
重心和对称性的稳定三原则。

1. 《粘粘世界》中的学习性分析

此次分析主要关注学习性与教育目标的吻合程度,即受试者对稳定三原则的理解和应用。因此,我们在游戏外的游戏前后测验中使用了增加用户行为表现的方法,并使用系统采集数据分析学习性是否可测量。在观察游戏内前后关卡的受试者通过率时显示:通过率小,但有明显改观。这可能与受试者游戏前进行教学关卡训练有关。因此,我们进一步测试了游戏内是否有特殊原则的性能加成,以获得反映用户所学知识的更精准的结果。也就是说,从前测到后测,用户都要在搭建过程中表现出基底宽、重心低和对称良好、结构更稳定的三原则意识。为此我们使用回放系统,计算用户每一关的状态指标,包括塔的基底宽度、塔重心位置和对称性测量(即从底部中心到重心和地心射线形成的角度),如图3所示。然后对比计算出所有用户同一关卡的相关得分与数值。结果显示:底宽得分与关卡的最大宽度相关,重心位置得分与关卡的最低位置相关,对称性得分与具有良好对称的90度角相关。结合游戏前与后的平均值得出结论:底宽和对称性原则在游戏后有所改善。即在游戏最后,受试者们开始注意考虑搭建较宽底边和对称性好的布局,但在重心位置意识中却没有改观,也就是说受试者根本没有意识到建筑中重心低更稳定的学习知识点。这说明设计师运用ECD分析法设计教育目标时,没有考虑到给用户足够的提示证明重心知识的重要性,因此导致用户忽略了该知识点。

2. 《粘粘世界》中的对齐分析

特伦特在研究游戏学习性时表明了游戏目标与教育目标对齐的重要性,通过研究发现,游戏可以适当刺激有目的的行为模式。^[12]因此,《粘粘世界》也研究了此问题,并评估了对齐性问题对关卡设计,甚至是游戏核心机制设计的重要性。另外,通过游戏还检测了“稳定性指标(stability metric)”。从现实重力学的角度来看,稳定性指标与建筑节点上水分子的聚集数量有关。水分子在分支节点上聚集越多,活动越频繁,建筑就越不稳定;而稳定值越高,建筑越稳定。使用数据分析对齐性,目的是通过指标反映建筑与用户通过某一特定关卡响应的程度。通过分析发现:底宽、重心高度、对称性和稳定性的指标与成功通过关卡存在必然联系。但是,重心指标却指向了错误方向。也



3. 指标示意图

就是说,该游戏中的重心原则设计与教育目标背道而驰,用户搭建重心较高的建筑反而可能通过关卡。设计师在设计重心原则关卡时脱离了宽底原则,给用户造成重心高、窄底也可以过关的误区。因此,对齐性与学习性分析结果一致,都可以找到设计缺陷并及时补救。另外,游戏难度、游戏时长、操作步数与可玩性问题也值得研究。我们通过粗略分析发现,《粘粘世界》可以循序渐进地引导玩家搭建简单建筑时如何处理底宽与对称性原则。

用户行为表现方面在每种分析法中都有细微的差别,学习性分析要求我们观察用户通过关卡的最后一种解决方案,对齐性分析要求我们关注游戏自身设计并将用户解决方法组成一个分析数据库。如果没有回放分析引擎(RAE),这些问题都得在庞大而昂贵的研究中完成,同时还要面对复杂的用户行为特征。目前,这些分析法只适用于单人物理教育游戏,无法处理大型三维多用户的游戏,如具有多用户虚拟环境的社交游戏《虚拟人生》。若使用上述分析法运载这类大型游戏需要收集更多日志数据,避免每个用户每分钟移动产生的日志数据量超负荷运行。一种替代性方法是只有在用户进入或离开特定区域时才记录日志数据,最终产生的数据会像“热度图”分析法一样。

结语

物理教育游戏有较大的分析难度,需要通过观察用户行为、记录日志数据完成多重分析。传统的日志记录分析法需要收集大量研究数据作为问题和测量法的指标,而新分析法只需在

工具包和C语言环境下的Unity引擎中便可完成。此外,不能只把某一个原则作为通关标准,应该综合稳定三原则进行整合性分析。由于研究精力和研究时间限制,本文的研究尚有诸多不足之处,希望其他研究者能够在本文基础上深化教育游戏学习性测量方法研究,并规范测量方法和分析技术,提高教育游戏设计领域的方法论研究。

*基金项目:本文为中国海洋大学青年教师专项基金“海洋文化创意型资源的数字化开发研究”(3014000-841613008)的阶段性研究成果。

注释:

- [1] Honey, M.A. and Hilton, M. *Learning Science Through Computer Games and Simulations*. National Academic Press, Washington, D.C, 2011.
- [2] Young, M.F., Slota, S., Cutter, A.B., et al. Our Princess Is in Another Castle: A Review of Trends in Serious Gaming for Education. *Review of Educational Research*, 2012, PP.61-89.
- [3] Erik Harpstead, Brad A. Myers and Vincent Aleven. *In Search of Learning: Facilitating Data Analysis in Educational Games*. ACM New York, NY, USA 2013, PP.79-80.
- [4] Spring, F. and Pellegrino, J.W. *The Challenge of Assessing Learning in Open Games: HORTUS as a Case Study*. Proc. GLS 2011, PP.200-208.
- [5] Gee, J.P. *What video games have to teach us about learning and literacy*. Palgrave Macmillan, New York, 2003.
- [6] Andersen, E., Rourke, E.O., Liu, Y., et al. *The Impact of Tutorials on Games of Varying Complexity*. Proc. CHI 2012, PP.59-68.
- [7] Delacruz G C, Chung G, Baker E L. *Validity evidence for games as assessment environments*. 9th International Conference of the Learning Sciences, Chicago. 2010.
- [8] Andersen, E., Rourke, E.O., Liu, Y., et al. *The Impact of Tutorials on Games of Varying Complexity*. Proc. CHI 2012, PP.59-68.
- [9] Corbett, A.T, Anderson, J.R. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction* 4, 1995, PP.253-278.
- [10] Shute V J. Stealth Assessment in Computer-based Games to Support Learning. *Computer Games & Instruction*, 2011. PP.503-524.
- [11] Koedinger K R, Cunningham K, Skogsholm A, et al. *A Data Repository for the EDM Community: The PSLC DataShop*. In, 2010. PP.43-55.
- [12] Torrente J, Freire M, et al. Evaluation of semi-automatically generated accessible interfaces for educational games. *Computers & Education*, 2015, 83: PP.103-117.