

教育游戏评价指标的设计与开发

刘文辉¹ 王艺亭¹ 赵敏¹ 胡贺宁¹ 江丰光^{1,2}

(1. 北京师范大学 教育技术学院 北京 100875;
2. 北京师范大学 未来教育高精尖创新中心 北京 100875)

[摘要] 对中小学理科教育游戏的合理评价,有助于使用者选择合适的教育游戏,促进游戏开发者开发出更加优质的游戏作品。然而目前针对教育游戏评价指标的研究缺乏。因此,本研究首先对已有教育游戏的评价框架进行综述,在分析6款目前最流行的中小学理科教育游戏,并对8名学生进行访谈的基础上,融合教学设计理论、心流理论(FLOW)与软件开发理论,建立了中小学理科教育游戏的评价指标;其次,邀请国内18位相关领域专家,通过德尔菲法对指标进行了三轮修改,利用层次分析法确定了评价指标的权重;最后,提出的评价指标体系包含学习设计、游戏设计、软件开发3个一级指标、11个二级指标、41个三级指标,并对每个三级指标评判的内容进行了详细的描述,可为学生、教师挑选教育游戏以及游戏开发者开发游戏提供借鉴。

[关键词] 中小学理科;教育游戏;评价指标;德尔菲法

[中图分类号] G434 **[文献标识码]** A **[文章编号]** 1007-2179(2017)02-0111-10

一、引言

随着现代科学技术的发展,移动设备、多媒体应用等越来越多地渗透到中小学课堂中,一线老师开始关注高效、多元、有趣、易懂的学习方式。教育游戏是教育信息化时代产生的一种寓教于乐的教学方式,它以教育为目的,以游戏为手段,融知识性、娱乐性于一体,对中小学生的学习和生活发挥了重要作用。近年的地平线报告亦提到“基于游戏的学习(Game-Based Learning)”为教育发展的新方向。

教育游戏的发展伴随着不同类型游戏的出现,同时也在不同学习情形中扮演不同的角色。例如,帮助学

习数学的游戏“龙箱(dragon box)”将数学与生活情境结合,通过循环的动机激发、探索、反馈、应用的学习过程,帮助学生自定步调学习数学(DragonBox.com, 2016);物理学题材的益智游戏“粘粘世界(World of Goo)”利用粘性小球构建建筑框架,提高学生对建筑结构的兴趣。此外,还有对历史学习有帮助的历史模拟游戏“刺客信条(Assassin's Creed)”、“文明(Civilization)”以及帮助语言学习的游戏“异能(Xenos)”、“动物捉迷藏(Peekaboo Barn)”等(Sawyer, 2014)。然而,市场上的教育游戏越来越多,很多教师、家长、学生面临教育游戏的选择难题。

回顾教育游戏的相关文献,我们发现评估使用教

[收稿日期] 2017-02-07 **[修回日期]** 2017-02-20 **[DOI 编码]** 10.13966/j.cnki.kfjyyj.2017.02.012

[基金项目] 北京师范大学教育学部2014年科研业务费专项资金资助(CXTD201401)。

[作者简介] 刘文辉,硕士研究生,北京师范大学教育技术学院,研究方向:教学设计;王艺亭,硕士研究生,北京师范大学教育技术学院,研究方向:绩效技术;赵敏,硕士研究生,北京师范大学教育技术学院,研究方向:技术增强语言学习;胡贺宁,硕士研究生,北京师范大学教育技术学院,研究方向:在线协作学习;江丰光,副教授,硕士生导师,北京师范大学教育技术学院,研究方向:学习空间、学习科学、STEM教育(fkchiang@bnu.edu.cn)。

[致谢] 感谢参加本研究填写调查问卷的所有专家;感谢接受访谈的8位中小学学生,及所有对本研究提供帮助的人。

育游戏学习效果的研究较多,这类研究将团队开发的或已有的教育游戏应用在教学中,通过实验或准实验研究,设置对比组,探索教育游戏对学生高阶思维、创造力、合作能力、情感等的影响(冀津等 2013; Juan & Chao 2015)。而有关教育游戏本身的评价研究比较少,如张文兰等(2011)、唐丽洁(2015)都采用内容分析法对教育游戏国内外文献进行分析,提出使用软件的效果研究居首位,而评价研究较少。因此,对教育游戏的评价研究有待进一步拓展(沈娟等 2014)。一套全面性的评价体系不但可以帮助使用者挑选符合需求的的游戏,也可以为游戏开发者提供参照,以便开发出具有教育性和游戏性的教育游戏。

二、教育游戏评价综述

对于教育游戏评价指标维度如何划分这一问题,不同学者有不同的看法。范云欢和崔金英(2008)通过对国内外已有网络教育游戏评价量表的分析,制定了包含游戏性、教育性、技术性三个维度,以及情景设置和情景丰富性等十五个因素的评价量规。叶长青等(2009)根据布鲁姆的目标分类理论,提出了数字化教学游戏评价体系的三维框架,包括知识、认知过程及游戏性。南京师范大学教育游戏研究中心从情境、内容选择、学习氛围、反馈判断、挑战机会、晋级机会、评分七个角度对游戏型软件进行分析,提出了基于教育视角的电子游戏分级分类思路(任秀平等 2009)。刘旨峰和林俊宏(Liu & Lin 2009)采用内容分析法对市场现有的 196 款计算机教育游戏进行了分析,结合专家意见提出评价指标,然后运用德尔菲法进行修改验证,最后提出的评价体系包括五个维度 43 个指标,五个维度分别为游戏基本信息、多媒体特征、游戏交互与结构设计、游戏内容和游戏反馈机制。在这些研究中,不同的学者选择不同的理论基础,构建了不同的评价指标。有的侧重于对教育内容质量的考察,有的侧重于对游戏可玩性的关注。虽然不同学者采用了不同的理论基础,建立的评价指标体系也各不相同,但他们对教育游戏的评价,都是从游戏的可玩性、教育目标的实现以及游戏的技术质量三个方面衡量的。这为我们构建中小学理科教育游戏评价指标提供了逻辑框架。然而,如何在教学内容与游戏主体之间寻求平衡,已有研究没给出良好的解决办法,仍是一个

亟待解决的问题。

除国内学者外,许多国外学者对影响教育游戏使用效果的因素以及教育游戏的评价进行了研究。有些研究通过实证方法探究了对教育游戏产生影响的因素,如卞在焕(Jaehwan Byun, 2015)测试了画外音对 18-25 岁孩子心流体验的影响,作者设置控制组和对照组,用一款游戏进行测试,最后得出游戏画外音对游戏沉浸体验有促进作用,游戏具备画外音会使游戏沉浸体验效果更显著。关于教育游戏评价的研究,学界主要聚焦在评价方法的探究及评价指标的制定上。在评价方法方面,奥马尔和加菲尔(Omar & Jaafar 2010)提出了启发式可玩性评估,这是一种使用一套相对简单、通用、有启发性的规则对教育游戏进行可用性评估的方式,主要过程包括:观察者解释评估对象;评估者了解或使用评估对象;评估者评估;集体讨论。在评价指标方面,唐迪和莫雷蒂(Dondi & Moretti 2007)关注游戏在终身学习中的应用,详细分析和展示了帮助教育者、实践者和终身学习者进行教育游戏质量评判的两个欧洲项目研究成果,其中的指标体系包括教学性与情境、内容、技术三个维度 62 个指标。评价指标较全面,且针对一定的使用对象。贾利勒(Djelil 2014)指出,在学习环境的研究中研究者需要考虑可用性、有用性及可接受性三个方面,也是对于其评价需要考虑主要内容。可以看出,国外教育游戏评价的研究较为成熟,但是对于 K12 阶段特定学科的研究较少。

对建立教育游戏评价框架的研究进行分析,可以为本研究提供借鉴,同时可以揭示出已有研究存在的问题,如评价指标针对性不强等。不同年级和不同学科的教育游戏在侧重点上有很大不同,比如理科游戏更多地培养学生的数理能力、逻辑思维能力、空间能力。文科类教育游戏侧重于人文素养的体现。然而,现有的关于教育游戏评价指标的研究中,缺少针对某一领域、某一年龄段的研究。

透过国内外文献综述,可以看出教育游戏的指标研究多数是基于某一理论构建的,如多元智能理论、目标分类理论等,然而,每一种理论都有其它理论无法代替的优越之处,仅仅运用单一的理论对教育游戏进行评价,是片面的。此外,目前的评价指标体系不够全面,缺乏详细的指标描述及权重。在实际操作过程中,评测人员根据自己对标准的理解,对

游戏内容进行审核,可能具有一定的主观臆断性。针对以上问题,本研究以中小学理科教育游戏为研究主体,制定出操作性强针对性强的教育游戏评价指标体系。

三、研究设计

本研究分为三个阶段:前期调研、德尔菲专家论证、AHP(Analytic Hierarchy Process,简称AHP,层次分析过程)权重确定。前期调研主要是对教育游戏评价研究以及市场上的教育游戏有清晰的认识,要收集和分析相关文献,并试用和分析一些教育游戏。本研究的对象是中小学教育游戏,所以前期调研中还包括对中小学游戏使用者的访谈,以了解用户对教育游戏关注的重心。前期调研结束后,我们对初步构建的指标体系在综合专家意见的基础上进行修改并通过层次分析法确定评价指标的权重。

(一) 评价指标的制定

本研究首先对已有教育游戏的评价指标进行了分析、梳理,在借鉴以前的研究成果和对中小学生的访谈结果以及排名较高具有代表性的教育游戏案例分析的基础上,确定评价指标。

1. 教育游戏的案例分析

目前,市场上的教育游戏种类繁多,各具特色。我们选取了不同平台(包括pc版、移动版)、不同学科(数学、物理、化学、生物)、不同风格的六款游戏:速算六边形、蜡笔物理学、沙盘玩具、数学速算、化学小子之危险中的帕蒂克(Chem Caper-Petticles in Peril)、细胞争霸(Cell Craft)。这几款游戏在相应游戏网站排名较高,有代表性且有媒体专家推荐。

本研究主要从教育性(针对的学科、知识技能、教育目的)、游戏性(游戏设定、情境、趣味性、激励机制)、技术性(界面设计、操作过程、音乐、游戏体验等)三方面进行了详细的调研,保障指标的完备性。

2. 学生访谈

为了兼顾学习者的建议,本研究在制定指标前对8位中小學生(包括2名小学生,5名初中生,1名高中生)进行了电话访谈,了解他们对教育游戏的认识及看法,为评价指标的开发提供参考。

访谈问题包括:

1) 访谈者基本信息。

2) 到目前为止,你用过哪些教育游戏?

3) 在这些软件中,你印象最深的是哪一款?为什么?

4) 什么样的教育游戏最吸引你?构想一个你喜欢的教育游戏?

5) 你觉得从教育游戏中学习对你是否有帮助?如果有帮助,体现在哪些方面;如果没有,原因是什么?

(二) 评价指标验证

基于以上研究,本研究设计出包括3个一级指标、12个二级指标、47个三级指标的教育游戏评价指标体系。为了论证评价指标的合理性,采用德尔菲法向18位专家进行了三轮咨询。

德尔菲法关键优点是避免了专家的直接对抗,将专家小组的讨论过程结构化,使他们在讨论复杂问题时更加有效(Okoli & Pawlowski 2004),不仅可用于预测领域,而且可广泛用于各种评价指标体系的建立和具体指标的确定过程(胡春萍等 2007;田军等 2004)。本研究在建立评价指标体系之后,应用德尔菲方法收集专家意见,对指标进行增加、删除、修改。具体来说,德尔菲方法的流程包括:明确评估目标、确定专家、发布问题、专家对问题的评估、对获取的专家知识进行处理(刘伟涛等 2011)。这一过程可以被概括为专家选取、问题咨询、数据处理三个步骤。

1. 专家选取

德尔菲方法的使用,需要根据所需要的知识范围确定专家,一般在8-20人左右(徐国祥 2005)。针对中小学理科的教育游戏,为了考虑不同专家的建议,本研究选取了教育游戏、教学设计、软件开发、中小学教师四个面的专家进行咨询。

根据研究需要,我们均选取具有一定研究工作经验的,对此领域比较熟悉的研究者或工作者,最后选取的18位专家中,包括4位研究教育游戏专家,4位教学设计专家,7位软件开发专家,以及3位中小学教师。其中教授2名,副教授7名,博士2名,10年以上高级教师2名,且有4名工程师和1名产品经理(见表一)。

在第一轮专家咨询中,我们将不同领域专家对二级指标的熟悉程度进行统计,熟悉程度分值从1到5,分值越高,熟悉程度越高。专家的熟悉系数表

现出不同方面的专家对各项的熟悉程度不一样,总体均值最低为 3.39,最高为 4.56(见表二)。可以看出,专家对教育游戏不同方面的熟悉程度总体较高,有较好的代表性。

表一 专家从事的专业领域

专业领域	人数	职称与学历	百分比(%)
教育游戏	4	2名教授 2名副教授	22
教学设计	4	1名教授 3名副教授	22
软件开发	7	1名博士,1名副教授,5名程序员(3-5年工作经验)	39
中小学教师	3	1名博士(9年数学教学经验),1名生物老师(10年以上高级),1名数学老师(10年以上高级)	17

表二 专家对二级指标的熟悉程度

二级指标	熟悉程度
交互设计	4.56
软件质量	4.39
信息管理	4.39
媒体质量	4.33
合作性	4.28
故事性	4.22
动机激发	4.22
学习内容	3.83
学习策略	3.82
学习评价	3.78
学习目标	3.72
学习者	3.39

2. 三轮专家咨询

根据德尔菲法,我们对选定的专家进行了三轮问卷咨询,问卷通过电子邮件及纸质发放。第一轮问卷内容包括专家对二级指标的熟悉程度,对三级指标的重要性(1-5)打分,以及认为需要增加或删除的指标、其他问题及建议等。我们根据第一轮专家的打分及建议,对指标进行了增加、删除及修改。第二轮问卷除要求专家对评价指标进行再次打分外,包括第一轮专家打分的分布、平均值、方差以及修改建议。第二轮调查后我们对指标进行了再次修改和第三轮调查。三轮咨询后,专家意见趋同。

3. 数据处理

本研究利用 SPSS 对回收数据进行处理。专家对指标重要性打分的平均值、方差、满分比,代表专

家意见的集中程度。变异系数和协调系数代表专家意见的协调程度。第 j 个指标的变异系数为 $V_j = \frac{\delta_j}{\bar{X}_j}$, δ_j 表示第 j 个指标的标准差; \bar{X}_j 表示第 j 指标的均数,说明专家对第 j 指标相对重要性的波动程度, V_j 越小,表明专家的协调程度越高。协调系数反应不同专家意见的一致性,范围在 0-1 之间,数值越大表示专家协调程度越好。

(三) 权重的确定

通过德尔菲法确定评价指标之后,本研究采用层次分析法确定指标权重;设计了一级指标和二级指标的 AHP 问卷,向教育游戏相关专家发送,最后回收有效问卷 5 份,通过数据分析,确定了二级指标的权重。由于三级指标的数目较多,利用 AHP 问卷不易收集数据且一致性检验可能无法通过,因此通过第三轮专家对三级指标的重要性打分,确定三级指标的权重。

四、研究结果

(一) 访谈结果

对学生进行访谈有助于了解他们对教育游戏的认识及观点(见表三),为指标设计提供参考。

1. 学生对教育游戏概念模糊

在访谈学生中,只有一名学生明确提出“什么是教育游戏?”问题,大部分学生并不清楚教育游戏是什么。他们列举的教育游戏,有魔抓(Scratch)等学习软件,也有课件。教育游戏是能够培养游戏使用者的知识、技能、智力、情感、态度、价值观,并具有一定教育意义的计算机游戏类软件(吕森林,2004),也就是说,教育游戏应该具有游戏特性和教育功能(魏婷等,2009),游戏是知识的载体和表现方式。课件或学习软件没有游戏性,只是单纯的学习工具,不能称为教育游戏。所以,对教育游戏进行清晰界定非常必要。

2. 学生对教育游戏的目的认识不同

部分学生认为教育游戏并无作用,是打发无聊时间的工具;部分学生认为是一种锻炼思维、想象力等的工具;还有部分学生认为有助于背诵记忆。出现这些认识,可能跟学生所谓的教育游戏多为象棋、围棋、学习软件等相关。一方面说明教育游戏可能在中小学使用率不高,另一方面说明目前市场上的

表三 参与访谈学生基本信息

访谈对象	访谈者信息	性别	年级	印象最深的教育游戏	最关注教育游戏的哪些方面	教育游戏提供的帮助
访谈 1		女	高三	数独, 金山打字, 手术	技术性, 难度适当	无帮助
访谈 2		男	初三	弈城围棋	有意义, 有趣, 可联机, 有积分榜	开阔眼界, 放松
访谈 3		男	初一	Scratch mblock	简单容易上手, 有创意	锻炼想象力
访谈 4		女	初一	五子棋	可自选模式, 可以在线共同游戏	锻炼思维
访谈 5		男	初一	象棋, 2048 成语接龙	画面清晰, 新颖, 刺激	打发无聊时间
访谈 6		女	三年级	成语游戏和教材配套的数学游戏	科目较为丰富, 难度较高, 颜色丰富	帮助背诵记忆
访谈 7		男	三年级	唐诗 300 首, 国际象棋	不同难度, 有中英文, 有关卡	帮助学习象棋
访谈 8		女	初三	数独, 七巧板	多结果, 趣味性强	开拓思维

教育游戏在教育性方面比较薄弱,包括教育目的的制定、内容策略的选择等。这要求我们在评价指标设计时要考虑游戏能否达到教育目标。

3. 学生最关注的游戏特质是“好玩”

学生眼中的好玩其实是教育游戏的趣味性。8 位学生中,提到趣味性的学生有 3 名。游戏在情境设计、交互设计以及背景故事设计等方面都可以让学生觉得有趣,愿意投入其中。另外,游戏的难度也会决定其“好玩”的程度,有些游戏设计过于简单,学生很快感到无聊,所以访谈中有 3 名学生提到难度,希望难度适当或者多级难度设置,可以进行难度选择。

4. 学生关注教育游戏的创新性和新颖性

对于中小學生来说,新奇的事物具有独特的吸引力。传统的答题或者计算类游戏,学生容易失去兴趣,所以游戏的创新程度应是评价的重要维度。

(二) 数据分析

1. 专家的代表性和积极性

在第一轮专家咨询中,问卷全部收回。第二、三轮分别回收 15、11 份问卷。对应三轮咨询,专家的积极系数分别为 100%、83.33%、73.33%,说明专家关心本次研究,回收率基本满足研究的需要(见表四)。

表四 专家咨询积极系数

轮次	第一轮	第二轮	第三轮
问卷回收数(份)	18	15	11
专家积极系数(%)	100	83.33	73.33

2. 专家意见集中程度

本研究采用界值法,对小于均值一个标准差的三级指标进行重点考虑;根据专家意见对这些指标进行修改及删除。经过三轮咨询,均值和满分比都有所增加(见表五),均值从 4.17 上升到 4.42,由此可见,从总体上说专家对两次修改后的指标给予了肯定。

表五 专家评分集中程度

轮次	项目	均值	标准差	界值
第一轮	算数平均值	4.17	0.32	3.85
	满分比	0.41	0.16	0.25
第二轮	算数平均值	4.33	0.28	4.05
	满分比	0.60	0.13	0.47
第三轮	算数平均值	4.42	0.23	4.19
	满分比	0.46	0.20	0.26

3. 专家协调系数

根据常保成等(Pao-Cheng Chang et al., 2002)提出的标准,当变异系数 $CV \leq 0.3$ 时,表示专家意见达到高度一致;当 $0.3 \leq CV \leq 0.5$ 时,表示专家意见在可接受范围内;当 $CV > 0.5$ 时必须解释其原因。本研究三轮咨询的变异系数均没有超过 0.5,且第一轮变异系数范围为 0.09-0.34,大于 0.3 的有三个指标;第二轮变异系数范围为 0.09-0.26,且变异系数大于 0.2 的指标只有 3 个;第三轮变异系数均小于 0.2。这说明专家意见逐步趋同,评分更加合理,结果更为真实可靠。三轮咨询中专家的协调系数都有所增加,最后达到 0.203,且 p 值都小于 0.05,说明专家评分有一致性(见表六)。

表六 专家评分协调系数

轮次	ω	卡方	p
第一轮	0.18	149.246	0.00
第二轮	0.192	129.678	0.00
第三轮	0.203	89.277	0.00

4. 不同领域专家评分差异

在第一轮专家咨询中,学习内容维度下的科学性和组织编排、学习评价维度下的一致性、媒体维度下的声音和动画五项指标具有显著性差异, p 值小于 0.05,其它 42 项指标无显著性差异。在四个专家领域中,一线教师对五项指标的打分均值最高,软件开发专家对五项指标的打分均值最低(见表七)。

不同领域专家对指标重要性的打分在很多指标上表现出差异性,原因可能是不同领域专家对指标评分的理解不同,软件开发者可能从难度上评分,认为声音、组织编排等开发难度较低,而教师从对学生学习的重要性评分,认为这些指标对学习者的学习很重要。在这种情况下,我们对具有差异性的五个指标进行了更加详细的描述,让不同领域专家有一致的理解。

5. 权重的确定

通过 AHP 权重问卷的咨询,我们最终确定五份有效问卷,五位专家判断矩阵的 CR 值均小于 0.1,具有满意的一致性,且五位专家均为高校老师,具有一定的代表性。由于回收问卷数量较少,所以确定的权重只是作为参考。根据不同专家的均值,确定一级指标学习设计、游戏设计、软件开发的权重分别为 55%、29%、16%(见表八、表九)。

(三) 评价体系

根据三轮咨询问卷的数据处理结果及专家提出的建议,我们对初步建立的评价指标进行了修正,得到最终的评价指标体系,由学习设计、游戏设计、软件开发 3 个一级指标、11 个二级指标、41 个三级指标组成,并对每个三级指标评判的内容进行了详细描述,保证使用者可以理解。

1. 学习设计

教育游戏首先必须具备一定的教育性。教育游戏在设计时需根据学生的心理发展规律,进行相应的学习设计以保障其教育性。本研究根据教学设计的基本流程,观察教育游戏在主要环节中产生的作

表七 不同领域专家评分差异

专家领域	教育游戏		教学设计		软件开发		一线老师		F	p
	均值	标准差	均值	标准差	均值	标准差	均值	标准差	F	p
科学性	4.25	0.5	5	0	4	0	5	0	22.99	0*
组织编排	4	0.82	4.75	0.5	3.86	0.69	5	0	3.33	0.05*
一致性	4.5	0.58	4.5	0.58	3.57	0.54	4.67	0.58	4.39	0.02*
声音	4.5	0.58	4.5	0.58	3.86	0.38	5	0	5.13	0.01*
动画	4.5	0.58	4.5	0.58	4	0	5	0	5.32	0.01*

表八 不同专家权重打分

一级指标	二级指标	专家 1(%)	专家 2(%)	专家 3(%)	专家 4(%)	专家 5(%)	均值(%)
学习设计 (55.18%)	a1	8.60	19.96	4.40	3.17	13.80	9.99
	a2	15.33	3.99	15.95	6.17	3.69	9.03
	a3	14.13	5.51	30.83	20.72	2.96	14.83
	a4	12.30	5.51	12.80	22.47	7.14	12.04
	a5	12.30	10.49	7.45	10.49	5.73	9.29
游戏设计 (28.58%)	b1	6.99	34.09	3.57	5.46	5.56	11.13
	b2	20.98	11.36	10.71	16.38	27.78	17.44
软件开发 (16.24%)	c1	2.93	5.10	1.03	1.38	2.38	2.56
	c2	2.93	0.72	3.81	8.05	2.61	3.62
	c3	2.93	0.72	3.81	3.39	10.37	4.24
	c4	0.59	2.55	5.64	2.32	17.97	5.81

表九 权重回收问卷一致性

	CI	RI	CR
专家 1	0.04	0.58	0.07
	0.07	1.12	0.06
	0.00	0.90	0.00
专家 2	0.00	0.58	0.00
	0.09	1.12	0.08
	0.08	0.90	0.09
专家 3	0.00	0.58	0.00
	0.10	1.12	0.09
	0.06	0.90	0.07
专家 4	0.05	0.58	0.09
	0.05	1.12	0.05
	0.05	0.90	0.06
专家 5	0.00	0.58	0.00
	0.06	1.12	0.05
	0.05	0.90	0.06

用,从学习者、学习目标、学习内容、学习策略、学习评价五个方面评价教育游戏的学习设计。

在教育游戏设计的过程中,设计者需要考虑其针对学习者的初始能力、一般特征,保证教育游戏符合学生的能力水平和心理发展特征。与其他阶段的教育游戏相比,中小学阶段更注重学习者的一般特征;学习目标的设定必须明确、具有一定的针对性,与学习者通过游戏达成的目标一致;对于理科教育游戏来说,学习内容必须是准确、科学的。此外,因为理科知识之间有一定的逻辑性,游戏在进行整体内容编排时,也要有一定逻辑,保证内容的广度和深度;理科教育游戏的学习策略,要选择合适的游戏类型,其中学习行为应采用实验探究等活动,且具有一定的创新性,在学习者遇到困难时,可能需要有帮助解决问题的工具,而不是直接提示或给出答案,更注重与学习者及时反馈的交互。在设计评价时,理科教育游戏的目标不是单一维度的记忆等,游戏中问题的解决办法可能有多种,评价的作用一方面是评判教育目标的达成,更重要的是帮助学习者反思与总结。基于此,评价会显得比较重要且贯穿整个游戏过程,需要提供新颖合适的评价活动或者挖掘过程数据,评价后反馈的内容也要能激发学生的探究欲。学习设计维度指标如表十所示。

2. 游戏设计

对于一款游戏来说,使玩家沉浸于其中,是最高

的游戏体验。只有学生投入游戏中,并带来愉悦感,才可能在游戏过程中实现相应的学习目标。席克真特米哈依(Csikszentmihalyi, 1975)将心流(flow)定义为一种将个人精神力完全投注在某种活动上的感觉,心流产生时同时会有高度的兴奋及充实感。根据米哈里的心流体验理论,让学生产生心流体验的因素主要是游戏中存在的挑战和学生掌握的技能,当挑战与技巧程度都很高时,可以达到较好的心流体验。

理科教育游戏的设计,主要从游戏的故事情境及激发动机设置评价。游戏的故事性,需要在真实的情境中进行,情节设定具有逻辑,且故事情感积极向上。产生并维持学习者的学习动机,除了必要的激励机制之外,根据心流体验理论,游戏任务的难度需要与学生的技巧达到一定的平衡,任务之间相关联,且有合理公正的升级规则。游戏设计维度指标如表十一所示。

3. 软件开发

教育游戏作为一种应用软件,需要对它的开发维度进行评价,这方面理科教育游戏与其他教育软件是相似的。教育软件一般从软件质量、交互设计、媒体质量、信息管理四个方面进行评价。软件质量包括教育游戏的安装难度、稳健程度以及运行的流畅度;交互设计包括游戏画面设计的美观性、组件分布的合理性、学生在游戏中的控制程度以及操

表十 学习设计维度指标

学习设计 (55%)	学习者 (10%)	知识储备	3%	教育游戏与其针对的学习者的知识储备相匹配
		技术水平	3%	教育游戏的操作简单,考虑到其针对学习者的软件操作水平
		一般特征	4%	游戏考虑到其针对学习者的年龄特征、思维特征、心理特征等,如游戏中场景、角色的设计、语言的应用符合学习者的年龄特征,不采用不合适学生的暴力性游戏;任务的设计与学习者的抽象逻辑思维发展水平相符等
	学习目标 (9%)	明确性	3%	明确教育游戏的学习目标,学习者可以知道此次学习的结果或标准
		针对性	3%	与中小学理科的某个知识点相联系,针对性强
		定义目标与学习者达到目标的一致性	3%	教育游戏的学习目标是学习者通过游戏学习的过程可以达到的,整体游戏的设计与目标是一致的
	学习内容 (15%)	科学性	5%	内容准确与科学,及时更新
		组织编排	5%	内容的组织编排合理,具有一定的逻辑体系
		与学习目标的一致性	5%	学习内容的广度和深度能确保学习目标的实现
	学习策略 (12%)	脚手架	2.2%	针对学习者在玩游戏的过程中存在的困难,提供适当的脚手架,或者有适当的认知工具辅助
		媒体应用	2.4%	在游戏中呈现学习内容使用的媒体,如图片、动画等,符合心理学原理,与学习者特点、学习目标、学习内容一致
		学习活动	2.4%	在游戏中学习活动的设计符合合理科学学习的特点,多采用实验探究等活动,具有一定的创新性
		游戏类型	2.3%	根据学习主题选择合适的游戏类型:如角色扮演、模拟、动作、冒险等
		反馈	2.7%	在学习过程及时与学习者进行交互,告知学习者学习效果、学习时间、理解的正确与否等
	学习评价 (9%)	过程性	1.6%	在游戏过程中及最后安排适量且分布合理的评估活动
		类型适合	1.8%	游戏中评价活动的类型合适,具有一定创新性
		评估后的反馈	2%	评估后提供及时恰当形式的反馈,不同的反馈形式有不同的特点,如只告知学习者对错状态,让学习者在游戏中继续探索正确答案,维持学习者的学习动力;或是反馈给学习者容易理解且能让学习者认识到自己问题的答案及解释,帮助学习者在下一次游戏中更好的表现
		一致性	1.8%	评价内容与学习内容一致,可以很好地评价学习者是否达到目标
		反思性	1.8%	评价活动起到帮助学习者反思的作用,如评价的内容可以让学习者回顾反思自己的学习过程;反馈的结果不仅包括总结性评价结果,还包括对学习者的整个学习过程的记录及分析

表十一 游戏设计维度指标

游戏设计 (29%)	故事性 (11%)	故事逻辑	3.5%	故事的情节设定没有明显逻辑错误,符合基本规律
		情境化	3.8%	故事情境设定与生活实际联系紧密,具有真实性
		情感渲染	3.7%	故事情感积极向上,对学习者的学习态度有积极作用
	动机激发 (18%)	任务难度	4%	游戏中任务的难度设置合理,给学习者一定难度的挑战,又不至于太难,不同任务之间难度有递增趋势
		任务关联	4%	任务与任务间存在一定的联系
		任务升级	5%	游戏有一定合理公正的升级规则,维持学习者的学习动力
		激励机制	5%	设定多种奖励形式,例如排行榜、积分榜等;及时的反馈机制,激发学习者学习动机等

作的难易性;媒体质量为游戏中包含的文字、图像、声音、视频的质量,以及它们之间的一致性;信息管理主要是对学生游戏进度的保存、游戏过程的完整跟踪以及个人信息数据的隐私保护。这几个方面是对教育游戏的技术性要求,也是游戏使学生达到学习效果的必要设计。本研究四方面权重分别为3%、4%、4%、6%,信息管理的权重较高,可见,对于理科游戏来说,数据的保存,尤其是学习过程数据的保存也是一项重要指标。软件开发维度指标如表十

二所示。

五、总结与展望

游戏在人们的生活中越来越重要,除了娱乐性,不能忽视的是它有助于激发学生学习潜能(Sawyer, 2014)。游戏如何更好地与教育融合是学习科学家们一直关注的问题。由于游戏开发的低门槛,导致市场上的教育游戏纷繁众多且良莠不齐,不合格的游戏很难吸引学生的兴趣,促进他们的学习。对教

表十二 软件开发维度指标

软件开发 (16%)	软件质量 (2.4%)	安装难度	0.8%	安装简单,尽量不需要在安装软件的同时安装其他运行环境;对电脑配置的要求低,保证大部分电脑可以使用
		软件稳健程度	0.8%	软件运行稳定,不轻易报错与崩溃
		流畅度	0.8%	游戏运行过程中,画面、声音、动画流畅,不会出现卡顿现象
	交互设计 (3.6%)	清晰美观	0.9%	画面简洁、清晰、美观,能够吸引学生注意力,具有较好的视觉体验
		组件排布	0.9%	画面中的组件摆放合理,符合学习者习惯,便于操作
		游戏控制	0.9%	发挥学习者的主体选择性,学习者可以自己选择游戏环境,例如游戏背景、角色;控制游戏进程,例如开始游戏、暂停游戏、退出游戏等
		操作难度	0.9%	操作过程简单,且有帮助新玩家熟悉操作的演示、提示等
	媒体质量 (4%)	文字	0.8%	文字清晰,重点部分有突出显示
		图像	0.8%	图像清晰美观,呈现合理,
		声音	0.8%	声音清晰,音质较好
		视频动画	0.8%	视频动画清晰,播放流畅
		文字、图片、声音和视频动画的同步性	0.8%	指音乐、文字与画面的内容显示一致,节奏相同
	信息管理 (6%)	进度保存	2%	可以保存学习者的游戏进度
		学习跟踪	2%	利用录屏或数据记录或其他技术,记录学习者使用教育游戏的学习过程,供学习者回顾与自我评价
		个人数据的隐私保护及安全	2%	数据保密性好,不会出现丢失及被盗的现象

育游戏的合理评价可以帮助学习者找到满足需求的教育游戏。而教育游戏的评价研究还存在一定的局限,如针对特定领域教育游戏的评价研究相对缺乏;在评价方法上,多为定性研究,缺乏定量实证研究;评价指标的开发多从教育者的角度,而非从学生的角度等。

作为一次探索,本研究通过德尔菲法,考虑 K12 阶段学习者与理科学科的特征,建立评价指标体系,并对每个指标进行了详细的描述及权重的确定。在使用时,评价者不具有专业知识也能理解,具有较高的可操作性。我们提出的指标体系基于教学设计的理论框架,而游戏作为学生学习、教师教学的一种策略,按照教学设计流程进行设计,可以将内容与游戏形式融合,而不仅是整合在一起,将不适合游戏学习的内容进行游戏设计。

当然本研究也有一定的局限性。由于条件有限,访谈对象数量较少且只有学生;三轮调查没能全部回收专家的问卷,专家的积极系数虽然可以达到要求,但相对较低;没有将提出的指标体系进一步试用分析等。这些缺点,希望可以在以后的研究中得以解决。针对指标体系的进一步试用、修正,开发出量表也是本研究下一步的工作。

[参考文献]

- [1] Byun, J. H., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments[J]. *Computers in Human Behavior*, 46: 129-138.
- [2] Chang, P. C., Tsou, N. T., Yuan, B. J. C., & Huang, C. C. (2002). Development trends in taiwan's opto-electronics industry [J]. *Technovation*, 22(3): 161-173.
- [3] Csikszentmihalyi, M. (1975). Play and intrinsic rewards [J]. *Journal of Humanistic Psychology*, 15(3): 135-153.
- [4] Djelil, F., Sanchez, E., Albouy-Kissi, B., Lavest, J. M., & Albouy-Kissi, A. (2014). Towards a learning game evaluation methodology in a training context: A literature review [A]. Busch, C. (2014). *Proceedings of the European Conference on Games-Based Learning* [C]. Berlin, Germany: Academic Conferences International Limited: 676-682.
- [5] Dondi, C., & Moretti, M. (2007). A methodological proposal for learning games selection and quality assessment [J]. *British Journal of Educational Technology*, 38(3): 502-512.
- [6] DragonBox. com (2016). *Apps in the Classroom* [EB/OL]. <http://dragonbox.com/educators>
- [7] 范云欢, 崔金英 (2008). 网络教育游戏评价量规的开发与应用研究 [J]. *中国教育信息化*, (6): 10-12.
- [8] 胡春萍, 杨君 (2007). 德尔菲法在构建政府绩效指标体系中的应用——以乡镇政府为例 [J]. *陕西行政学院学报*, 21(4): 12-15.
- [9] Juan, Y. K., & Chao, T. W. (2015). Game-based learning for green building education [J]. *Sustainability*, 7(5): 5592-5608.
- [10] Liu, E. Z. F., & Lin, C. H. (2009). Developing evalu-

ative indicators for educational computer games [J]. *British Journal of Educational Technology*, 40(1): 174-178.

[11] 刘伟涛, 顾鸿, 李春洪(2011). 基于德尔菲法的专家评估方法[J]. *计算机工程*, (S1): 189-191.

[12] 吕森林(2004). 教育游戏产业研究报告[J]. *中国远程教育*, (22): 44-47.

[13] Okoli, C., & Pawlowski, S. D. (2004). The delphi method as a research tool: An example, design considerations and applications [J]. *Information Management*, 42(1): 15-29.

[14] Omar, H. M., & Jaafar, A. (2010). Conceptual framework for a heuristics based methodology for interface evaluation of educational games [J]. *Computer and Information Science*, 3(2): 211.

[15] 任秀平, 李艺(2009). 电子游戏的分级与分类问题的教育视角论证[J]. *远程教育杂志*, (2): 66-69.

[16] Sawyer, R. K. (2014). *The Cambridge handbook of the learning sciences* [M]. Cambridge: Cambridge University Press Page: 413-416

[17] 沈娟, 章苏静(2014). 教育游戏评价方法研究述评[J].

远程教育杂志, (3): 105-112.

[18] 唐丽洁(2015). 国内十年游戏化学习研究现状与分析[J]. *中国教育信息化*, (10): 23-25.

[19] 田军, 张朋柱, 王刊良, 汪应洛(2004). 基于德尔菲法的专家意见集成模型研究[J]. *系统工程理论与实践*, 24(1): 57-62.

[20] 魏婷, 李艺(2009). 国内外教育游戏设计研究综述[J]. *远程教育杂志*, 17(3): 67-70.

[21] 徐国祥(2005). *统计预测和决策* [M]. 上海: 上海财经大学出版社: 11.

[22] 叶长青, 王海燕, 王萍(2009). 数字化教学游戏三维评价体系架构[J]. *远程教育杂志*, 17(6): 71-73.

[23] 冀津, 王冠宇(2013). 悦趣化学习在面向对象程序设计教学中的应用研究[J]. *计算机教育*, (18): 99-102.

[24] 张文兰, 牟智佳(2011). 悦趣化学习的研究现状与热点分析——基于国内外期刊数据库论文的分析[J]. *现代教育技术*, 21(6): 32-38.

(编辑: 李学书)

Developing Evaluative Indicators for the Educational Games

LIU Wenhui¹, WANG Yiting¹, ZHAO Min¹, HU Hening¹ & CHING Fengkuang^{1 2}

(1. School of Educational Technology, Beijing Normal University, Beijing 100875;

2. Beijing Advanced Innovation Center for Future Education, Beijing Normal University, Beijing 100875, China)

Abstract: Educational games are useful for K12 students to study science. Evaluation is helpful both for users to select proper educational games and for designers to develop high-quality games. However, there is a lack of research about developing indicators to evaluate educational games in science for K12. The purpose of this research is to build the evaluative indicators by reviewing evaluation research, analyzing six popular educational games, interviewing eight students in K12, and combining the theories of instructional design, FLOW theory and software development theory. Through a Delphi study approach, the indicators were refined by eighteen experts, and the weight was determined using Analytic Hierarchy Process. The indicator system of this study includes 3 first-class indicators (learning design, game design and software development), 11 second-class indicators, and 41 third-class indicators, which helps teachers and parents to select suitable games for their students and children and provides reference for developers.

Keywords: science of K12; educational games; evaluative indicators; Delphi study