

基于真实用户评论信息构建移动学习资源的评价模型

程罡¹,高辉¹,余胜泉²

(1. 国家开放大学 北京 100085; 2. 北京师范大学 北京 100875)

【摘要】移动学习资源的开发和应用已经成为 E-Learning 的重要发展方向,而目前对于移动学习资源的系统化评价方法仍然缺乏成熟的模型。在已有相关研究的基础上,作者利用内容分析的研究方法,分析 App Store 中 40 个排名靠前的学习资源类应用程序的 1000 条真实用户的评论信息,来构建移动学习资源的评价模型。通过对这些评价信息进行内容分析和归类,统计各个指标项下的评论信息的出现次数来确定评价模型中各指标项的权重,并具体描述了各指标项下典型的用户评论内容,据此提出设计移动学习资源时在这些方面应当注意哪些要点。与传统的领域专家构建资源评价指标的方式不同,该研究提供了一种以用户视角出发的一种评价框架,对于移动学习资源的设计者、开发者,以及选择、购买移动学习资源的个人和机构用户,都有一定的参考和研究价值。

【关键词】移动学习资源,内容分析法,评价模型,评价量表

【中图分类号】PT391

【文献标识码】B

【文章编号】1001-8700(2014)01-0043-11

一、研究背景

随着智能手机、平板电脑的快速普及,基于这些智能设备的交互式学习资源开始成为一类重要的数字化学习资源,并被越来越多的学习者所接受。以最大的移动应用商店——苹果 App Store 为例“截止到 2013 年 4 月,App Store 中的教育类应用达到 89000 多个,占活跃应用总量的 11% 以上,是仅次于游戏的第二大应用类别”^[1]。这些新型的学习资源大量利用了移动设备上的触摸操作方式、强大的多媒体功能,表现出较强的互动性、趣味性,受到相当一部分学习者的喜爱,也体现出许多与传统的数字化学习资源不同的特点。

对于一些传统的数字化学习资源,已经有很多成熟的评价模型、工具或量表来描述和评估资源的质量或者对特定需求的匹配程度。例如学习对象的评价模型和量表^[2-4]表,主题学习网站的评价量表^[5-6],网络课程的评价模型和标准等^[7-8]。这些评价模型和量表不但为用户评估和选择学习资源提供了可量化的方法,同时也为资源的设计和开发者

提供了原则上的参考,能够指导他们在资源开发过程中遵循基本的设计原则和规律,避免一些常见的失误和原则性的错误。

显而易见,相对于主题网站、网络课程等传统学习资源,移动学习资源有其自身的特点和适用性,用户在选择移动学习资源时也有不同的视角和价值取向,因此应当设计针对移动学习资源的评价模型和工具来帮助用户更好的选择适切的资源。目前很多移动应用商店提供了简单的用户评分和评论的评价方式,从某种程度上可以满足一部分资源筛选和评价的功能,但是其精细化程度还远远达不到用户的实际需要。例如,无法快速判断资源的内容质量、数量、操作体验、难易程度等。

因此,我们认为,在目前移动学习逐渐成为一种普及化的学习方式的背景下,移动学习资源的使用者需要更全面、精细化的移动学习资源评价模型,来帮助他们更好的搜索、选择和评估移动学习资源。同时,资源的设计和开发者可以借助该模型更好地指导资源开发,而资源的集成、发布者(一般是移动应用程序商店)也可以提供更好的评价调查表而非

【基金项目】北京市教育科学规划课题(编号 CJA10244)“基于‘学习元’的远程教育泛在学习资源共建共享和应用研究”的研究成果。

【作者简介】程罡,国家开放大学教育学院讲师;高辉,国家开放大学信息管理处实习研究员,硕士;余胜泉,北京师范大学教育学部教授,博士生导师。

简单的评分,获取更细化的资源评价信息。

二、相关研究

1. 学习对象的评价工具

学习对象(Learning Object)是数字化学习资源的一个重要类型,它的许多核心理念如“可重用性”、“自足内聚”、“目标指向”、“可聚合”等^[9],都成为后来数字化学习资源设计与开发的必备特征。同样,许多移动学习资源在某种程度上也是学习对象这一理念在移动学习环境下的实现和进一步发展。因此,适用于学习对象的资源评价模型可以作为移动学习资源评价模型的重要参考,其中Nesbit等人提出的LORI(Learning Object Review Instrument)是比较经典的学习对象评价模型。在LORI的最新版本中,对于学习对象主要从以下9个维度进行评价^[10]:

表1 LORI 1.5版本的指标和主要内容描述

评价指标	主要内容描述
内容质量	真实性、准确性、合适的内容表达、详略得当
学习目标的一致性	学习目标、活动、评价和学习者特征的一致性
反馈和适应性	根据不同的学习者模型和输入提供适应性的内容或反馈
动机激发	能够激发学习动机,调动学习兴趣
表征设计	设计合适的视觉和听觉信息来促进学习
交互设计	导航便捷,用户界面符合心理预期,能够获取有效的帮助
可获取性	设计合适的控制功能、界面格式使之适应残疾人、移动终端的使用需要
可重用性	能够在不同的学习情境下使用,适应不同的学生背景
标准兼容性	符合相关的国际标准或规范

观察以上表格不难发现,主题学习网站、网络课程等其他类型学习资源的评价模型很多都是在此基础上发展和衍生出来的,其核心的一级指标同样也适用于移动学习资源的评价。因此,LORI模型可以作为构建移动学习资源评价模型的基础,但是我们仍然需要针对移动终端的特性、移动学习者的使用情境和习惯,调整部分指标项中的具体内容。

2. 移动学习资源的评价模型和量表

国内较早针对移动学习资源的质量开展专门研究的是詹青龙在2009年提出的移动学习资源质量评估模型^[11],其中针对内容型的学习资源,提出了

5个方面的质量维度:

- (1) “内容符合移动学习目标,编排尊重学科规律,学习小模块化、微型化”;
- (2) “综合运用多种媒体素材增强学习趣味性”;
- (3) “对各种素材和数据信息的描述和组织便于入库、调用和改进”;
- (4) “软件界面设计风格简洁、统一,各要素布局合理,易识别,导航定位方便准确”;
- (5) “能提供关键词跳转标记功能,有操作帮助和反馈系统,如专家提醒、FAQ等”。

我们不难发现该质量评价模型和LORI模型还是非常相似的,只是在评价指标的具体描述中通过强调内容的微型化、界面风格的简洁等因素来突出移动学习的特点。

陈明明等提出的移动学习资源评价量表中从内容、结构和技术三个方面对学习资源进行评价,但是相对于LORI模型而言,并没有提出新的评价指标项,只是对于“链接的外观名明显”、“书签定位”等细节提出了一些新的要求^[12]。由此可见早期的移动学习资源评价模型并没有脱离LORI评价模型的范畴,只是进行了一些局部的调整。这也反应出当时移动学习资源的整体特征实际上与传统的数字化学习资源区别不大。

而近几年来,由于智能触屏移动终端的迅猛发展,出现了大量基于这些智能终端的移动学习App(基于智能手机的应用软件的简称)。针对这些App,国外的一些机构和研究者开始尝试提出专门的评价模型和量表。

例如芝加哥公立学校图书馆部门的“Mobiliary”项目组在2011年设计了评估iPad上的App是否适合学校图书馆使用的量表,该量表从以下三个方面来构建评价指标^[13]:

表2 Mobiliary设计的iPad App评价量表

评价指标	分项内容描述
对学习的支持	支持特定的学习目标;阅读材料符合目标群体的年龄;
	内容呈现符合教学的原则;内容的准确性与权威性;
	能够提升高阶思维能力;能够促进协作;
	内置了测评功能,能够追踪学习过程、保留学习记录;
	内容符合国家/州立的教学标准;

易用性	导航和菜单易于使用; 有使用帮助; 包含支持服务的页面;
	清晰地呈现了教学目标; 对不同年龄阶段设计了不同的应用层次;
	内容分享; 多媒体控制开关; 使用图形增强理解; 支持多种语言;
质量	能够扩展数字图书馆的资源; 文字表达恰当;
	能够提升学生的参与度;
	价格合理或者免费; 不包含广告; 能够长期使用;

值得注意的是,“Mobiliary”提出的评价量表中

表3 Villar的移动学习App评价量表

评价指标	描述	和LORI指标项的对应关系
内容	内容的准确性、恰当的表达、与学习目标的一致性、内容的聚焦。	基本上与LORI中“内容质量”、“学习目标”两项指标对应,但是Villar在此处着重强调了移动学习资源“聚焦”的重要性,即提供的学习内容贵精不贵多。
个性化	能够根据不同的情况调整学习内容和交互活动。	与LORI中的“反馈与适应性”指标相对应。
反馈	能够适时的提供足够的反馈,帮助学习者自我检测。	
高阶思维能力	能够训练学习者的高阶思维能力而非简单的知识记忆。	这是LORI中没有出现过的指标项。
易用性和技术性能	学习者易于掌握,操作失误后能够快速返回,一段时间不使用后再次使用是否容易,在移动设备上运行的速度快慢等等。	与LORI中的“交互设计”、“可获取性”相对应。
交互性和沉浸性	人机交互设计是否合理,创设的环境和交互是否有吸引力。这项指标	基本上与LORI中的“交互设计”、“表征设计”、“动机激发”相对应,但此处更强调通过设计有趣的、与真实情境结合的交互活动,来增强学习体验,促使学习者沉浸其中,这也反映了当前移动学习资源非常注重用户体验的特征。
社会交互	学习者在使用该App时与其他学习者能否方便的交流,分享信息和学习成果。	这是LORI中没有出现过的指标项。

与之类似的评价模型还有Vincent的移动学习App评价量表^[15],Schrock的基于内容的移动学习App评价量表^[16],eSkillsLearning网站的移动学习App评价量表^[17]等等,其具体内容与Villar的评价模型大同小异,在此不再赘述。

3. 研究现状评述

总体而言,目前的移动学习资源评价模型的研究还停留在比较初级的阶段。首先,这一领域还缺乏有广泛影响力和被普遍应用的成熟模型,目前已有的模型大多是根据研究者个人经验构造,缺乏足够的实证研究和数据支持,因此各个指标项的权重没有量化的指标;第三,这些模型大多从研究者自身的视角出发,而缺乏最终用户对移动学习资源的真实评价信息。在应用这些模型进行评价时,这都是至关重要的问题。我们认为,目前移动App市场已

的二级评价指标项,开始出现一些LORI指标体系中未曾提及的特征,例如“提升高阶思维能力”、“促进协作”、“追踪学习过程和保留学习记录”、“内容分享”、“价格与广告”等。

另一个比较具有代表性的例子是教学设计师Villar提出的评价标准,她综合了自己设计e-Learning教学的经验以及对iPad上一些移动学习App的案例研究经验,建议从7个方面来评估移动学习App^[14],如下表所示:

经有海量的用户评价信息,这些能够成为构建更完善的移动学习资源评价模型的重要数据基础。我们希望在现有评价模型的基础上,充分利用移动学习资源的终端用户提供的真实评价信息,来构建更贴近用户需要的移动学习资源评价模型。

三、研究设计

1. 研究方法

本研究尝试以客观的传播内容为依据来构建一个抽象的分类模型,内容分析法是一种非常合适的研究方法。所谓内容分析法,是“用客观系统并量化的描述显性传播内容的一种研究方法”^[18],它的一般操作方式是多个研究人员对选定的显性传播内容(文本、视频、音频)等进行内容分析,然后根据一定的规则将这些内容进行赋值、量化处理和统计分

析(最典型的例子是将某个时间区段内特定主题的新闻进行内容分析,通过量化的统计各个分类中的新闻数量来观察媒体对于某一特定事件的态度),最后通过这种分析手段来作为某种观点、理论的佐证,或者作为某个研究的探索式前期研究的组成部分。克林格(Kerlinger)认为^[19]：“内容分析法在概念上类似于态度测量中使用‘观察法’,近似于观察人们的行为或‘要求他们填答量表’”。但是内容分析法的优点在于其非干扰性,传播内容的提供者本身不会在事先就知道他们提供的信息会成为某个研究的研究对象,因此很好的保证了其客观性。另一方面,相对于社会科学的其他质性研究方法,内容分析法通过编码小组的量化和测量,可以将大量的信息简化为数字和模型,能够更好的反映这些信息之间的内在联系。

通常来说,内容分析法的一般步骤包括^[20]:(1)提出研究问题;(2)确定研究范围;(3)抽样;(4)选择分析单位;(5)建立分析类目;(6)建立量化系统;(7)进行内容编码;(8)分析数据资料;(9)解释结论”。其中内容编码环节一般需要对编码员进行训练,并进行实验性研究检验编码员的信度,以保障研究的可信度和可重复性。

2. 基于内容分析法的移动学习资源评价模型构建过程

我们认为,内容分析法对于我们的研究问题而言是一个非常有效的方法。参照前文所述我们将研究问题确定为:从用户的视角来考察移动学习资源,他们会从哪些方面来评价资源质量?这些方面各自所占的权重如何?由此我们可以确定一个质量评价模型的基本结构。

目前主流的移动应用程序商店中已经积累了非常多的真实用户使用后的评论信息。这些是非常有效的传播信息:对于大多数用户而言,只有那些给他们留下深刻印象的软件才会使他们有动力上去发表评论,而这些评论信息无疑能够有效地表明他们最为关注的特性——无论是优点还是缺点。通过有效的抽样并获取相当数量的用户评价信息进行内容分析,并将它们归类到不同的评价指标维度,能够分析出从我们构建的评价指标是否能够反映用户所关心的特征,并且用量化统计的方式可以了解不同评价指标的重要性。同时,由于移动应用程序往往是兼容多个操作系统的,因此只需要选择一个主流的应用程序商店,其研究样本就具备很好地扩展性。所

以我们的研究范围确定为一定时间段内某个主流应用程序商店中的移动学习资源评论内容(具体抽样方法参见下文)。基于内容分析法的一般研究过程,本文构建移动学习资源评价模型的研究过程如下图所示:

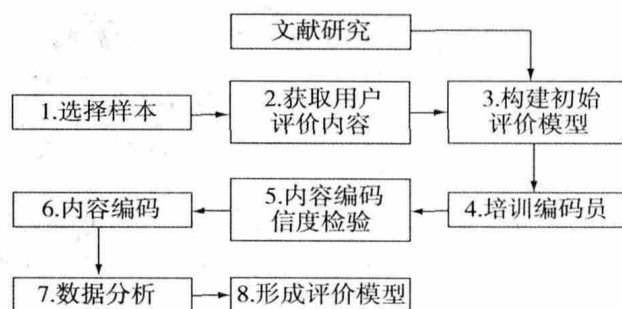


图1 研究过程

其中主要的工作包括:

(1) 选择一定数量的样本,并抓取用户评价内容。

(2) 根据已有的研究文献以及对部分用户评价样本的分析,构建一个初始的评价模型。这个模型作为以后将抓取到的评价内容进行分类的基础。

(3) 研读一定数量的评价内容,确定内容分析的协议,并培训编码员。

(4) 编码员对少量的评价内容进行编码,检验其信度。

(5) 信度符合要求后,正式对全部内容进行编码。

(6) 统计、分析内容编码的结果,完善评价模型,确定各评价指标项的权重。

四、研究过程

1. 初始评价模型构建

通过对已有研究文献的分析,我们不难发现移动学习资源的评价模型与学习对象的评价模型有很密切的联系,用于评价学习对象的很多指标项仍然适用于移动学习资源,例如内容的质量、教学目标的一致性、教学反馈、动机激励等,因为这些基本的指标反映了教学的基本要求,适用于任何类型的学习资源。

另一方面,在移动学习的情境下以上的指标项仍然需要进行优化和调整。根据我们对已有研究成果的研究,以及对随机抽取的部分用户评价信息的研读,我们从以下几个方面进行了调整:

(1) 内容维度拆分成内容质量和内容数量两个

方面。根据我们对用户评价信息的观察,很多用户评价信息中表明了对资源内容两个方面的关切:一方面是内容是否真实准确,是否易于理解等关于内容质量方面的信息;另一方面是关于内容的丰富程度、详略是否得当(避免内容过载),内容模块的大小是否适宜移动情境下的使用等关于内容数量的信息。考虑到在移动学习的情境下,学习者的时间、注意力、屏幕尺寸的局限性都比传统的数字学习环境下更多,学习资源的适量和充分聚焦成为非常重要的特征,我们有必要单独列出这方面的评价指标项。

(2) 学习目标的一致性和动机激发整合成“整体学习体验与一致性”。学习目标、动机激发更多的是从教育者的视角出发的一种描述方式,对于学习者而言,他们更关心的是学习资源在整体设计上是否能满足他们的需要,有一致的学习体验。因为移动学习者的时间和注意力极其有限,所以移动学习资源是否能创设优质的学习体验至关重要,包括迅速的引起学习者的兴趣,吸引学习者持续投入注意力,并且在学习过程中不容易受到干扰。我们认为这是一个对移动学习资源非常重要的指标项。

(3) 反馈与适应性拆分成两个维度:学习反馈与评价、情境适应和个性化。我们认为,一方面学习评价是学习设计中一个不可或缺的组成部分,有必要设计一个单独的一级指标来进行描述;另一方面情境性和适应性已经成为移动学习资源区别于传统学习资源的一个重要特征,因为移动设备能够更方便的获取学习者和周围环境的情境信息,情境适应性和个性化是未来移动学习的一个重要发展方向,因此我们将“情境适应和个性化”也独立成一个单独的一级指标项。

(4) 交互设计拆分成两个维度:人机交互和社会交互。传统的学习资源评价模型更多考虑的是人机交互的设计,如“导航便捷,用户界面符合心理预期,能够获取有效的帮助”^[21]。而当前大家越来越关注学习资源是否能促进学习者的社会交互,社会交互已经成为移动学习的重要特征之一,余胜泉等认为“以移动和普适计算为技术基础,以分布式认知、社会建构主义、情境认知为理论基础,以协同知识建构、泛在学习、社会认知网络构建为特征的分布式情境认知将成为一种重要的学习范式”^[21]。因此我们认为,移动学习资源是否具备促进学习者之间建构社会认知网络的功能,促进分布式认知,也应成为评估移动学习资源的一个重要维度。

(5) 可获取性的指标应考虑学习资源对不同人群、不同尺寸终端、不同网络环境的支持,同时,由于大量的移动学习资源以商业软件的形式存在于各类应用程序商店中,因此资源的价格也是衡量其可获取性的一项指标。另外还应考虑是否包含多语言的版本,让使用不同语言的学习者使用。

(6) 去掉“可重用性”和“标准兼容性”两项指标。我们发现,“可重用性”和“标准兼容性”两项指标在移动学习资源的各类评价模型中很少被提及,在用户提供的评价内容中也几乎没有出现。这也反映出目前移动学习资源还处于发展的初期阶段,绝大部分资源都是独立运行,鲜见多个资源相互包含、配合的案例,也缺乏成熟的资源重用技术标准。因此,在我们的初始模型中暂时不考虑这两个LORI中的指标项。

基于以上分析,我们构建的移动学习资源评价的初始模型如下表所示:

表4 移动学习资源评价的初始模型

评价指标	指标描述
1. 整体学习体验与一致性	整体设计上能给学习者带来愉悦、沉浸式的学习体验; 内容、评价和交互方式与学习目标、学习者特征保持一致;
2. 内容质量	内容的真实性、准确性、科学性; 内容可读性强、易于理解;
3. 内容数量与资源粒度	内容丰富、能够满足学习者的需要; 内容详略得当、模块大小适合移动学习情境下使用;
4. 学习反馈与评价	能够根据学习者的表现给予恰当的反馈和脚手架; 提供恰当的学习评价让学习者了解学习进度和效果,评价方式适合移动学习情境;
5. 媒体表征	对学习内容的多媒体呈现和表达具有较强的吸引力和感染力,能够激发和维持学习者的学习兴趣;
6. 人机交互	操作简单、易用; 导航清晰、人机界面设计合理,交互流畅、自然; 能够对资源内容的播放和呈现予以适当的控制,或附加书签、批注
7. 情境适应和个性化	能够根据周围的环境、学习者的状态变化给予适应性的学习内容和交互活动; 学习者能够调节难度、做个性化的定制;

8. 社会交互	学习者之间能够有效的沟通交流,建立和维系人际关系; 能够进行内容的分享与协作,分享学习成就;
9. 可获取性	没有明显的技术错误和使用障碍; 设计合适的控制功能、界面格式使之适应不同人群、不同终端和不同网络环境的需要; 资源的价格合适,学习者能够承受; 多语言版本;

2. 样本选取和数据抓取

移动学习资源的类型很多,例如移动的电子书(EBook)、移动学习网站、短信和彩信类资源、以移动应用程序(App)形式存在的资源等。其中App是种类最为多样化、学习体验和交互最好的一类移动学习资源,而且各个主流的移动操作系统都有完整的App商店,其中的用户评价信息都是公开的,比较容易获取这些信息进行内容分析。因此,我们选取App类的移动学习资源作为本研究样本选取的范畴。

目前市场上主要的移动智能操作系统主要包括iOS、Android、Windows Phone等,其中iOS系统的应用都集中在Apple App Store中,便于进行集中的检索和获取;同时iOS也是发展最早和最成熟的移动应用生态系统,其中的教学类资源非常丰富,也深受教师和学生的喜爱。许多优秀的教学类App都是首先开发iOS的版本,再迁移到其他移动操作系统中,因此我们考虑将Apple App Store中的教学资源类App作为本研究样本选取的来源。

App Store中的应用程序数量及其庞大,总数超过100万个,即使只考虑教育类的应用程序也有接近20万个。而这些应用程序中,下载量分布是非常不平衡的,下载排名1/1000的应用占据了全部下载量的一半以上,而有近一半的应用几乎无人问津。在这种背景下,我们认为使用概率抽样的方法选取样本更贴近研究的实际,即下载次数更多(使用人数更多)的应用应有更大的概率被抽中。由于无法获得app store中所有应用的精确下载次数,我们用另一种简化的方法代替:即使用app store教育分类排行榜中排名靠前的n个应用作为我们的抽样范围,因为他们是在概率上有非常大的可能性被用户下载和使用的。

我们在2013年3-4月,用中文和英文两种语言对Apple App Store的教育分类进行检索,并按照其热门程度进行排序(苹果提供排序结果)^[22-23],选择其中属于“学习资源”的App(剔除掉教育游

戏、工具类的教育应用)作为初步的分析对象。由于App Store中的热门App有很多是非常相似的产品(例如英语100句、法语100句,背单词软件等),为了突出样本的多样性,对于内容与形式比较类似的App我们只选用排序最靠前的一个。同时考虑到研究小组的人力所限,最终依据排序和筛选后的结果,抽取了20个英语类App,20个中文类App,最终入选抽样范围的App如下表所示:

表5 抽样入选研究范围的热门学习资源类APP列表

中文 App	英文 App
儿童必背诗词 50 首 HD	iTunes U
每日必听英语 美国口语 8000 句	Kids Can Match
经典儿歌	TeachMe: Kindergarten
网易公开课	Lumosity Brain Trainer
有声故事书 - 睡美人(中英双语)	Kno Textbooks
有声《三字经》注音(简体) 典藏版 Lite	Sight Words List - Learn to Read Flash Cards & Games
英语流利说	Alphabet Tracing
宝宝学 ABC HD	Survival Guide
找驾校 - 交规考试 2013 版	3D Brain
不可不知的 1000 个历史常识	TED
口头交际王	BrainPOP Jr. Movie of the Week
儿童学画	PBS KIDS Video
宝贝画画看 - 儿童绘画 临摹 互动猜图	Moon
宝宝学身体部位 HD	Spanish FREE 24/7 Language Learning
慢速 VOA 精华	ABCmouse. com Zoo Set 1
家居装修设计 - 你的家装好帮手	Craftsy
开心宝宝 - 迎春版	SkyView Free - Explore the Universe
幼儿园数学全套训练	History: Maps of World
沪江听力酷	Science360 for iPad
魔力小孩识字	Baby™

由于Apple App Store的每一个App对应于不同的国家和地区都有单独的用户评价页面,因此我们编写了一个脚本程序,可以抓取某个特定App在不同国家和地区App Store中的用户评价信息,并存入待分析的数据表中。在抓取了以上App所有的评价信息以后,根据评价内容的丰富程度排序,对每一个App选取前25条用户评价内容作为我们的研

究样本,每一条评价信息的字数从2000字到数十字不等。

3. 信度检验与内容编码

“内容分析中的信度可以定义为不同编码员对内容归类的一致性”^[24]。如果不同编码员对同一个内容可能编入不同的类目,那么人们的偏见就会以不可知的、无法控制的方式影响内容编码的结果,从而影响整个研究的可信度。一般而言,基于内容分析法的研究的信度通过较明晰的类目定义和描述、对编码员的训练以及对抽样内容进行内容编码信度检验等方式得到保障。

另外一个很重要信度指标是考量用户评论的真实性。由于移动App的商业属性,所以有部分厂商花钱购买“刷榜”服务,雇佣一些厂商使用大量的虚假账户发表正面评价以促进App销量。在各种应用程序商店中,App Store的管理和审核是最为严格的,并且苹果在2012年底到2013年严厉打击“刷榜”行为,2013年初以后所谓的“刷榜”服务几乎在App Store中销声匿迹,这在一定程度上保证了我们研究样本的可信度。此外,由于我们在众多评论信息中优先选取内容最多的评论进行分析,进一步减少了虚假评论信息的可能性,因为大部分虚假评论都是针对App的评分,而对评论文本没有特殊的要求,大部分都会使用一些简单、重复的文本评论内容,因此我们的样本选取策略能够很好地筛选掉这些虚假评论,保证研究的信度。

本研究所构建的评价指标分类是基于比较成熟的学习资源评价模型而构建的,编码员是长期从事该领域研究的专业人员,对此非常熟悉和了解,因此在类目定义方面是有较好的信度保障的。在研究过程中,参与编码的两位编码员使用独立编码、汇总果、讨论和修订不一致的编码的方法,进行了5轮的编码训练(100条评论内容样本),并且在此过程中不断完善类目的定义和描述,形成编码协议。

在经过训练后的独立编码检验中,两位编码员的编码一致性达到了80%以上,符合内容分析法信度检验对置信水平的一般要求^[25]。在此基础上,两位编码员对20个中文App和20个英文App的1000条用户评论进行内容分析和编码,去除掉过于简单、无法提供明确信息的评论后,共计得到1253个分类条目(一条评论信息可以被归类到多个分类维度),并对每一个分类条目的价值判断进行区分(正面评价信息为赞扬和肯定,负面评价信息为批

评和建议)。其中英文评论信息条目数为751条,中文评论信息条目数为502条。各个评价指标对应的简单数量比例如下表所示:

表7 内容编码结果统计表

评价指标	用户评价数量比例
1. 整体学习体验与一致性	总体比重: 17.88% ,其中正面评价的比例: 94.64%
2. 内容质量	总体比重: 13.95% ,其中正面评价的比例: 76.02%
3. 内容数量与资源粒度	总体比重: 10.45% ,其中正面评价的比例: 43.51%
4. 学习反馈与评价	总体比重: 4.39% ,其中正面评价的比例: 67.27%
5. 媒体表征	总体比重: 6.46% ,其中正面评价的比例: 69.14%
6. 人机交互	总体比重: 19.71% ,其中正面评价的比例: 50.2%
7. 情境适应和个性化	总体比重: 4.23% ,其中正面评价的比例: 62.26%
8. 社会交互	总体比重: 1.68% ,其中正面评价的比例: 66.67%
9. 可获取性	总体比重: 21.55% ,其中正面评价的比例: 51.48%

4. 对内容编码的分析

从数量分布上来看,9(可获取性)、6(人机交互)和1(整体学习体验与一致性)是用户最为关心的三项指标,其评价数量占据了总量的6成左右,平均每项占到了20%左右的权重。

2(内容质量)和3(内容数量与粒度大小)是第二档次的指标项,平均每项占到了12%左右的权重。

5(媒体表征)、4(学习反馈与评价)、7(情境适应和个性化)是第三档次的指标项,平均每项占到了5%左右的权重。

8(社会交互)几乎不被用户关注,总体比重不到2%。从对用户评价的内容分析来看,这个指标项的重要性不足以单独列出,可以与6(人机交互)合并为“交互设计”的评价指标项。

各个评价指标项的详细分析如下:

(1) “可获取性”

传统的资源评价模型中不太受关注的“可获取性”指标,在移动学习资源的用户评价中反而是最受关注的方面,这是移动学习资源的设计者特别需要注意的。这反映出移动学习资源在下载、使用等方面遇到的问题更多,要求也相对更高。值得注意

的是,即使是这些排名靠前的 App,这个评价指标项的正面评价比例也不高,只有 51.48%。从评论内容看,主要的问题集中在频繁闪退和报错、价格高昂、不同网络和终端下适应性不强等等。

并且我们在研究时发现,“可获取性”的负面评价信息与其他类型正面评价的共现率非常低(只出现了 5 次):这说明一旦移动学习资源在“可获取性”方面满意度不高,学习者几乎不会关注到它其他的优点,即使这个资源在其他方面设计得可能非常不错。由此可以看出良好的“可获取性”设计是维系学习者后续学习行为的重要前提条件,这也是学习资源开发者从传统应用情境迁移到移动情境首先需要关注的一个方面。

(2) “交互设计”

在我们最初设计的评价量表中,设定了“人机交互”和“人际交互”两个指标项,但是从内容分析的结果来看,绝大部分用户目前仍然更关注“人机交互”,即操作界面、导航等方面。值得注意的是,“人机交互”出现的频率很高,但是其负面评价比例同样较高,这说明从传统学习资源迁移到移动学习情境下时,对于人机交互的优化设计,既是学习者非常关心的要点,也是资源开发者需要重点研究和改进的方向。

“社会交互”这一指标项出现的频率不高,但是集中度特别高,主要在少数 2、3 个 App 的评论中出现,而且占据了它们的评论信息中很大的比重。这说明社会交互虽然受用户关注不多,也不是必需的特性,但是如果设计得当也会成为一个鲜明的特色,受到学习者的欢迎。从内容分析的结果来看,只要资源中设计了“社会交互”这方面的功能,就会受到好评,因此这属于开发者还没有太关注,但是学习者比较喜欢的一个方向。

(3) 优秀的整体设计是受用户欢迎的关键

在以上的 9 个分类指标中,“整体学习体验与一致性”是获得正面评价比例最高的(94.64%),说明这些排名靠前的移动学习资源在内容、媒体表征、交互等分类指标方面可能各有千秋,但是要成为受到欢迎的 App,必须在整体设计上做得比较优秀。从用户的评论内容来看,大部分的正面评价是“有趣”、“吸引人”、“适合某类人群的学习”这几类。说明在整体设计上突出目标的针对性和一致性,同时保持趣味性和良好的使用体验是成功的关键。

(4) 资源的内容

我们在分类指标中将资源内容相关的评价项分成了“内容质量”与“内容数量与粒度”两个维度,如果合并计算的话,涉及资源内容的评论数量是最多的(接近 1/4),这说明资源内容仍是用户评估移动学习资源质量最重要的基础性指标。

其中“内容质量”的正面评价比例为 76%,在所有的评价指标项中仅次于“整体学习体验与一致性”85%的 App 中都至少出现了一条关于内容质量的评论,这说明较高的内容质量也是成为受欢迎的移动学习资源的一个基础性指标。在关于内容质量的负面评价中,大部分是针对内容准确性的一些改进建议,还有一些是对资源内容针对性的质疑。

“内容数量与资源粒度”是正面评价比例最低的指标项,仅有 43.51%。考虑到我们抽样的都是排名靠前的 App,这一数据更足以说明,如何充分考虑资源数量和粒度大小的权衡,是一个值得移动学习资源开发者充分关注的问题。这些负面评价较多的集中在资源数量不足(多见于免费 App),视频或电子书体积太大难以下载,或者提供的资源太多太杂乱、超过学习者实际需要等。作为移动学习资源的开发者,应当尽可能的考虑合适的资源粒度,同时资源要比较精确的聚焦于既定目标,在此基础上提供较多的资源供学习者选择。

(5) 其他指标项

“学习反馈与评价”、“媒体表征”、“情境适应与个性化”三个指标项在总体数量上相对接近,都在 5% 左右,并且价信息分布不均匀,集中分布于少数几个 App 上,例如“学习评价与反馈”分类项中有 55 个条目,其中 13 个集中在“TeachMe - Kindergarten”这个 App 中,并且全部都是正面评价,而大量 App 则完全没有这方面的评价信息。另外两个指标项也有类似的现象。

以上这些数据表明,这几个指标项与前几个基础性评价指标不同,属于“锦上添花”的要素。优秀的 App 不一定要在这几个方面都有突出的表现,但是一旦在这些方面有较好的设计,会给用户留下很深刻的印象(表现为这几个指标项的正面评价集中分布于少数几个 App)。

根据我们分析的结果,将各个指标项根据出现频率重新排序,并列出其细化分析的结果如下表所示:

表 8 内容编码结果细化分析表

评价指标	出现频率	典型的用户评论	评论内容中体现的主要特征
9. 可获取性	21.55% 覆盖率: 100%	<p>典型正面评价: "免费"、"下载方便"、"价格合理"、"在多个终端下都可以使用"</p> <p>典型负面评价: "软件经常崩溃"、"网络状况不好时无法使用"、"价格过高"</p>	<ol style="list-style-type: none"> 1. 出现频率第 1, 覆盖率达到 100% (每一个 App 都有这方面的评论), 最受用户关注; 2. 正面评价比例较低, 值得开发者重点关注; 3. 负面评价与其他评价指标项的正面评价共同出现的频率极低, 容易成为后续使用的障碍, 是需要优先解决的第一个"门槛"。
6. 人机交互	19.71% 覆盖率: 87.5%	<p>典型正面评价: "易于使用"、"操作很智能"、"搜索和导航很方便" 以及对人机交互的各种细节设计的描述和赞扬</p> <p>典型负面评价: "导航不够清晰和简洁"、"观看和播放资源的控制方式不够灵活"、"广告太多影响观看和操作"</p>	<ol style="list-style-type: none"> 1. 出现频率第 2; 2. 负面评价比例第 2 高, 尚有很大改进空间, 值得开发者密切关注; 3. 适合移动终端的操作方式和导航设计是值得重点研究的方向, 使用户易于上手, 但是又设计一些高级的、智能的交互功能, 是很受用户欢迎的设计方式。
1. 整体学习体验与一致性	17.88% 覆盖率: 97.5%	<p>典型正面评价: "很有趣"、"很吸引人"、"特别适合某个方面的学习"、"完全符合我的预期"</p> <p>典型负面评价: "不适合某类人使用"、"看不出有什么作用"</p>	<ol style="list-style-type: none"> 1. 出现频率第 3, 受到非常多的关注; 2. 正面评价比例最高, 说明排名靠前的资源在这方面普遍受到认可; 3. 趣味性、吸引力、针对性是被用户提及最多的词语; 4. 如果目标定位不明、整体体验不佳, 很难成为受普遍欢迎的移动学习资源。
2. 内容质量	13.95% 覆盖率: 87.5%	<p>典型正面评价: "内容很好/很合适"、"有很多我不知道的知识"、"提供的信息很准确"、"发音标准"等</p> <p>典型负面评价: "内容不准确"、"内容有错误"</p>	<ol style="list-style-type: none"> 1. 出现频率第 4, 覆盖率达到 87.5%, 是一项基础性指标; 2. 正面评价比例第 2 高, 说明优质资源在这一项上普遍做的不错; 3. 值得关注的仍然是内容的准确性、适切性。
3. 内容数量与资源粒度	10.45% 覆盖率: 80%	<p>典型正面评价: "内容很全面"、"非常多的信息"、"资源种类丰富"</p> <p>典型负面评价: "内容太单薄"、"内容太多太杂"、"视频太大难以观看完"</p>	<ol style="list-style-type: none"> 1. 出现频率第 5; 2. 负面评价比例最高, 有很大的改进空间; 3. 资源数量不足或过多都容易引发学习者的不满; 4. 要避免直接将粒度太大的资源向移动终端迁移, 要做适当的切分和微型化处理。
5. 媒体表征	6.46% 覆盖率: 67.5%	<p>典型正面评价: "漂亮的图画"、"很好视频效果" 以及一些媒体设计方面的细节描述和赞扬</p> <p>典型负面评价: "过于花哨"、"文字太小无法看清"、"用视频表达内容效果会更好"</p>	<ol style="list-style-type: none"> 1. 出现频率第 6; 2. 不是必需的要素, 但设计特别精美的资源能够得到集中的好评; 3. 出现的问题大多是在小尺寸屏幕上显示和观看的遇到障碍, 或用户认为缺乏一些适宜的媒体类型(音频、视频等)。
4. 学习反馈与评价	4.39% 覆盖率: 40%	<p>典型正面评价: "有趣的的游戏和测验"、"多样化的测验方式"、"即时的反馈和奖励"</p> <p>典型负面评价: "练习题太简单/太难/太枯燥"、"提供的指导太少"</p>	<ol style="list-style-type: none"> 1. 出现频率第 7; 2. 不是必需的要素, 但设计得好的资源能够得到集中的好评; 3. 大多数学习者喜欢资源提供评价和反馈要素, 特别是评价与反馈设计得当可以极大的增加他们的兴趣, 给他们留下很深的印象; 负面评价主要集中在评价方式是否得当、反馈内容是否详细等方面。
7. 情境适应和个性化	4.23% 覆盖率: 45%	<p>典型的正面评价: "可以定制内容"、"可以动态调整难度"、"智能的内容推荐"、"适应不同的网络环境"</p> <p>典型的负面评价: "无法定制内容/调整难度"、"特定情境下找不到合适的内容"</p>	<ol style="list-style-type: none"> 1. 出现频率第 8; 2. 不是必需的要素, 但设计得好的资源能够得到集中的好评; 3. 大多数评论显示学习者喜欢可以自己调整学习内容、学习进度、学习难度, 智能的资源推荐也很受欢迎。

8. 社会交互	1.68% 覆盖率: 27.5%	典型的正面评价: "能够发布我的得分"、"能够看到我的排名"、"能够分享给我的社交媒体"、"能够与其他用户交流" 典型的负面评价: 希望能够增加上述的社交内容	1. 出现频率最低,覆盖率低,但集中分布; 2. 不是必需的要素,但设计得当可以成为非常重要的优势和特色。因为只要提供这项功能的资源就会受到好评,而这个指标项下得负面评价全部是用户提出诉求,希望增加此类功能; 3. "排名"、"分享"和"交流"是最主要的用户诉求。
---------	------------------	--	--

注:其中“覆盖率”x表示某一项指标在某个资源评论中至少出现一次的比例,它与总体频率作比较可以反映该类评论信息的集中程度。

除了频次分析和典型评论内容的归类描述外,我们还希望通过分析每一个资源中不同指标下的评论数量是否有强相关性,如果两两之间相关性太强说明这两个评价指标相互间影响很大,可能不适宜作为独立的评价维度,需要重新考虑和调整。另一方面,我们也希望分析各个指标项对于用户正面评论比例的影响。因此,我们对40个样本资源的9个评价指标下的评论数量和正面评价比例放在一个矩阵中,进行多变量的相关度分析(皮尔森检验),分析结果如下表所示:

表9 相关度矩阵

	值向量间的相关性					
	整体学习体验与一致性	内容质量	内容数量	评价与反馈	媒体表征	人机交互
整体学习体验与一致性	1.000	-.062	-.160	.345	.233	-.045
内容质量	-.062	1.000	.079	-.041	-.097	-.112
内容数量	-.160	.079	1.000	.054	.149	.089
评价与反馈	.345	-.041	.054	1.000	-.001	.089
媒体表征	.233	-.097	.149	-.001	1.000	.317
人机交互	-.045	-.112	.089	.089	.317	1.000
情境与适应性	-.004	-.273	-.122	.058	.243	.232
人际交互	-.196	-.016	.034	.118	.051	.168
可获取性	-.220	.069	.157	-.051	.040	.189
正面评价比例	.320	-.116	.051	.227	.000	-.179

表9 相关度矩阵(续)

	值向量间的相关性			
	情境与适应性	人际交互	可获取性	正面评价比例
整体学习体验与一致性	-.004	-.196	-.220	.320
内容质量	-.273	-.016	.069	-.116
内容数量	-.122	.034	.157	.051

评价与反馈	.058	.118	-.051	.227
媒体表征	.243	.051	.040	.000
人机交互	.232	.168	.189	-.179
情境与适应性	1.000	.007	-.036	.130
人际交互	.007	1.000	.189	.124
可获取性	-.036	.189	1.000	-.413
正面评价比例	.130	.124	-.413	1.000

根据统计学上的一般规律,相关度数值在0.7以上可以视为强相关,表中并没有任意两个变量之间的相关度在0.7以上,因此可以认为这9个评价指标的独立性相对较好,并没有太大的重叠和干扰。另外,与我们关心的“正面评价比例”相关度最高的是“可获取性”(负相关)和“整体学习体验与一致性”(正相关),说明这两项的确是提升用户满意度的最关键的指标。

六、结论与讨论

根据以上的研究和分析,我们建议构建的移动学习资源评价模型可以包含8个主要的评价指标项:(1)整体学习体验和一致性;(2)内容质量;(3)内容数量和粒度大小;(4)交互设计(包括人机交互和社会交互);(5)可获取性;(6)媒体表征;(7)学习评价与反馈;(8)情境适应与个性化。各个评价指标项的描述参见表4。

各个指标项的权重,可以参照表7的数量分布,根据对移动学习资源的实际需要进行设定。如果作为通用的评价量表,我们建议将1-5作为必选的评价指标项,1、2、3的权重可设置为12%-20%左右,4、5的权重可设置为18%-25%左右;6、7、8作为可选的额外加分指标项,根据实际情况设置每项占据5%左右的权重,总权重在10%-15%之间。

当然,从严格意义上来说,任何模型都是错误或者不全面的,因为它们都是对现实世界中研究对象的一种简化、抽象,而不能反映全部的客观事实。本文只是尝试从用户使用移动学习资源后提供的真实评价信息出发建构一个评价模型,使资源的设计和

开发者能更好的了解哪些方面是最终用户特别关心的、哪些是必须完成的基础工作、哪些是可能的加分项、哪些方面是要极力避免的;同时这个评价模型也可以作为移动资源的评估者(可以是机构,也可以是希望使用移动学习资源的用户)全面的评估移动学习资源质量的一个参考指标。

除了各个指标项及其权重的确定外,本研究也通过对用户评价信息的内容解读和数量分布分析发现了一些有趣的特征,例如用户对移动学习资源质量的关注度最为集中的反而不是资源内容的质量和媒体表现,而是可获取性、交互设计、整体学习体验等方面。这也反映出设计移动学习资源,首先要考虑用户易于获取、导航便捷易于操作、整体学习体验较好,否则即使有较好的内容也不一定会受到用户的认可。再比如,即使对排名靠前的App而言,内容数量与资源粒度、人机交互和可获取性也是大有提升空间的(正面评价比例较低),移动学习资源的研究者和开发者可以在这些方面下更大的功夫进行针对性的研究,等等。

我们期待更多的研究者充分利用互联网上大量的用户真实评价信息,利用多种研究方法进行分析,对移动学习资源的评估模型进行更多、更细化的深入研究。

【参考文献】

[1]148apps. Biz. App Store Stats Summary [EB/OL]. <http://148apps.biz/app-store-metrics/?mpage=catcount>, 2013.

[2]John Vargo, John C Nesbit, Karen Belfer, et al. Learning object evaluation: Computer-mediated collaboration and inter-rater reliability [J]. *International Journal of Computers and Applications*, 2003, 25(3): 198-205.

[3]John Nesbit, Karen Belfer, John Vargo. A convergent participation model for evaluation of learning objects [J]. *Canadian Journal of Learning and Technology*, 2002, 28, (3).

[4]Tracey L Leacock, John C Nesbit. A framework for evaluating the quality of multimedia learning resources [J]. *JOURNAL OF EDUCATIONAL TECHNOLOGY AND SOCIETY*, 2007, 10(2): 44.

[5]陈美玲,柳栋,武健.教育主题网站的评价量表[J].*中国教育学刊*, 2003, 9, 55-57.

[6]周梅群,吴华.教育网站评价指标体系设计之研究[J].*电化教育研究*, 2004, 8, 53-57.

[7]朱凌云,罗廷锦,余胜泉.网络课程评价[J].*开放教育研究*, 2002, 1(35): 22-28.

[8]CELTS-22,网络课程评价标准[S]. 2012.

[9]胡小勇,祝智庭.学习对象——网络教学技术的新理念[J].*电化教育研究*, 2002, (4): 1-9.

[10]J. C. Nesbit, K. Belfer, T. L. Leacock. LORI 1.5: Learning Object Review Instrument [EB/OL]. <http://www.elera.net/>, 2004.

[11]詹青龙,元梅竹.移动学习资源建设的质量屋模型构建[J].*中国电化教育*, 2009, (10): 51-56.

[12]陈明明.移动学习资源的创设与评价探究[J].*软件导刊(教育技术)*, 2010, (10): 88-89.

[13]Chicago Public Schools Dept. iPad App Assessment Rubric [EB/OL]. <http://mobiliary.wikispaces.com/apprubric> 2011.

[14]Mayra Aixa Villar. 7 Essential Criteria for Evaluating Mobile Educational Applications [EB/OL]. <http://mayraixavillar.wordpress.com/2012/12/06/7-essential-criteria-for-evaluating-mobile-educational-applications/> 2012.

[15]Tony Vincent. Educational App Evaluation Rubric [EB/OL]. <http://learninginhand.com/static/50eca855e4b0939ae8bb12d9/50ecb58ee4b0b16f176a9e7d/50ecb593e4b0b16f176aa974/1330908312793/Vincent-App-Rubric.pdf> 2012.

[16]Kathleen Schrock. CRITICAL EVALUATION OF A CONTENT-BASED MOBILE APP [EB/OL]. <http://kathyschrock.net/pdf/evalipad.pdf> 2011.

[17]Eskilllearning. Mobile Application Selection Rubric [EB/OL]. http://eskilllearning.net/uploads/Selection-Rubric_Scoring_Rubric.pdf 2011.

[18]Bernard Berelson. Content analysis in communication research [M]. New York: Free Press, 195: 18.

[19]Fred N Kerlinger. Foundations of Behavioral Research. [M]. New York: Holt, Rinehart and Winston, 1973: 525.

[20]李本乾.描述传播内容特征,检验传播研究假设[J].*当代传播*, 2000, (1): 47-51.

[21]余胜泉,程罡,董京峰.e-Learning新解:网络教学范式的转换[J].*远程教育杂志*, 2009, 19(3): 3-15.

[22]Apple Inc. App store 教育类应用热门榜单(中文) [EB/OL]. <https://itunes.apple.com/cn/genre/ios-jiao-yu/id6017?mt=8> 2013.

[23]Apple Inc. App store 教育类应用热门榜单(英文) [EB/OL]. <https://itunes.apple.com/us/genre/ios-education/id6017?mt=8> 2013.

[24]Daniel Riffe, Stephen Lacy, Frederick G. Fico. Analyzing Media Messages: Using Quantitative Content Analysis in Research, Second Edition [M]. Taylor & Francis, 2005: 124.

[25]Klaus Krippendorff. Reliability in content analysis [J]. *Human Communication Research*, 2004, 30(3): 411-433.

(本文责任编辑:闫兵)