

在线学习行为特点及其影响因素分析研究^{*}

魏顺平

(中央广播电视大学 现代远程教育研究所 北京 100031)

【摘要】 中央广播电视大学和全国不少网络教育学院的网上教学检查表明,学生在线学习时间投入普遍偏低。已有研究表明,网络资源和学习支持服务等两个外部因素是影响学生在线学习投入的重要因素。本研究以中央电大开放教育入门课程“开放教育学习指南”网络课程为例,以2010年春季学期江苏电大9369名新生学习该课程时所产生的登录数据、资源浏览数据和形成性考试数据为分析对象,采用数据挖掘的方法得出登录、资源浏览、作业、测试等在线学习行为的特点及其影响因素。

【关键词】 在线学习; 行为特点; 影响因素; 数据挖掘

【中图分类号】 G434

【文献标识码】 A

【文章编号】 1007-2179(2012)04-0081-10

一般而言,在线学习是指通过因特网进行学习与教学的活动。在本文中,在线学习指的是发生在网络教学平台上的学习,要比一般意义上的在线学习(以及与之近义的 e-Learning、网络化学习、数字化学习)的外延要窄。本研究所关注的在线学习行为特点描述及其影响因素分析主要基于学习者在教学平台上留下的各种行为记录以及学习者的基本信息。研究远程学习者的在线学习行为特点,并发现影响这些行为的关键因素,将有助于远程教师及管理人员了解当前学习者的总体学习状况,为他们开展相关决策,如优化教学平台和教学资源、改进学习支持服务提供有力参考。

目前,有关在线学习行为描述与影响因素分析的研究围绕在线学习行为描述、影响因素分析及相关监控软件开发应用等展开。这些研究采用文献研究、问卷调查、访谈、网络日志分析、内容分析等方法对学习者的学习行为、参与度以及影响参与度的因素进行研究。

从已有研究来看,不少研究通过分析教学平台网络日志来描述在线学习行为。例如,谢维奇(2004)采用数据挖掘方法对 Web 服务器日志和后台数据库进行挖掘,一方面为教师及时提供学员的访问次数、访问时间、教学资源的利用情况,另一方面为网站的管理者提供用户的浏览模式信息,分析使用者的访问路径,从而使教师对学员的学习过程能实施有效的监控,完成对学员的形成性考核,并完善网站设计。宋江春等(2005)提出使用挖掘技术分析学生的访问行为、频度、内容、停留时间等内容,得出学生访问行为和方式的一般模式,并以此为依据向学生提供全面的个性化教学支持服务。袁明等(2006)则以一所网络教育学院教学平台学习记录为挖掘对象,设置课程页面点击次数和上网人数两个参

数,结合时间和地域特点,对已上网人数、每天访问在线课程人数及其月平均值、访问量时段特点、各地区(以下称为站点)学习者到访比例和在线测试题完成情况等方面进行分析研究。有研究人员将关联规则挖掘技术用于从用户访问序列数据库的序列项中挖掘出相关的规则,揭示学习者访问一些内容的同时会访问哪些内容,借此找出具有相关内容的网页,从而更好地组织课程页面和推荐页面,尽可能缩短相关内容的分布距离,或提供便捷的路径指引(Ha et al., 2000; Wang, 2002)。Romero 等(2005)曾采用统计、可视化、分类、聚类以及关联规则挖掘等数据挖掘方法对 Moodle 平台中的各种数据进行分析,得到的分析结果可为学生了解自己的学习状况、教师监控学生的学习状况提供参考,并为教师的教学决策提供依据。上述研究成果可为本研究的开展提供挖掘内容和挖掘方法参考。

已有研究表明,在线学习行为的影响因素主要有网络资源和学习支持服务。赵丕元(2002)通过对2000名中央电大开放教育试点学生的问卷调查发现,影响学生参与网上学习的不利因素有:网上资源内容尚不够丰富、网页形式呆板,有些课程辅导缺乏交互性、网页内容更新较慢;学习支持服务不够便捷。龚志武(2004)指出,网络课程结构体系相对完整,学生网上学习时间和次数就会相应增长;师生网上交互率的高低与学生网上学习的时间和次数成正比,即交互率越高,上网次数越多,学习时间也越长。肖爱平等(2009)指出在线学习的影响因素具体包括网络环境、课程资源、网上交互以及学习者自身的原因。

此外,在线学习行为分析的实际应用就是建立在线学习行为监控系统(毛利杰等,2009;杨金来等,2008;邵艳丽等,

^{*} 基金项目:本文系2011-2012年度联校教育社科医学研究论文奖课题“远程学习者在线学习行为特点及其影响因素研究”(编号:JY11043)成果。

2008)。网络学习中的监控机制对于促进网络学习这种新型的学习方式的普及和发展、提高网络学习质量具有重要意义。其中 杨金来等(2008)所设计开发的监控系统涉及来访客户端信息统计、不同地域学习者在线分布情况、学习者在在线人数的实时统计、学习者个体网络学习行为的实时监控等。

综合上述已有研究,笔者认为,基于网络日志的学生行为分析还可以更加细致深入,可以从学习时间、学习活动、师生交互、学习资源、考试成绩等多个角度发现更多的学习行为特点;影响因素分析还可更加注重内在因素,即学生自身因素的分析。本研究将采用网络日志分析这一非反应性研究方法以及相关分析、差异检验、聚类分析、社会网络分析等数据挖掘方法,对在线学习的网络日志做出更为细致深入的分析,并探索在线学习行为的内在影响因素。

课程概述与研究样本选取

“开放教育学习指南”是中央广播电视大学开放教育专科和专科起点本科各专业学生的必修课,也是开放教育学生的入门课程。该课程配有文字教材和网络课程。基于网络课程的“开放教育学习指南”课程试点工作已在江苏电大、天津电大以及中央电大直属学院开展。由于是网络课程试点,所有学生均需登录电大在线平台(<http://www.openedu.com.cn/>)进行学习,这样保证了后面所获得的学生网上学习过程记录是完整的。本文选择“开放教育学习指南”网络课程的在线学习过程数据作为研究样本,这些数据包括学生在使用电大在线平台时产生的登录、资源浏览、学习体验等数据以及学生在使用形成性考试平台时产生的考试成绩数据。为了开展影响因素分析,研究者还从教务管理平台中采集了学生基本信息数据。最终,研究者从三个不同的平台中获得了江苏电大参加2010年春季“开放教育学习指南”网络课程试点的学生的基本信息数据、登录数据、资源浏览数据和形成性考试数据等,并导入数据库,形成四张数据表。这四张数据库的关联通过“学号”这一字段来建立。经联合查询,这四张数据表的学生数并不一致,为了能对所有学生做一个完整的学习过程分析,作者对这四张数据表进行交集运算,取得这四张数据表中共有的学生,共计9369名。本研究即以这9369名学生对应的基本信息、登录、资源浏览和考试成绩等四类数据作为分析对象,开展远程学习者在线学习行为特点及其影响因素研究。

参加该网络课程学习的学生分为本科、专科和“课程开放”三个专业层次。其中,“课程开放”属于一种以课程为单元的开放学习形式,即允许学习者注册单科学习课程。这三种专业层次的学生分布情况如表一所示。我们以5年为一个单位,将学生年龄分为若干段,然后统计得到各年龄段学生的数量分布,其结果如表二所示。学生集中在15~34岁之间,这四个年龄段的学生数占样本总数的83%。

表一 2010年春季江苏电大参加试点的学生专业层次分布

专业层次	男	女	总计
课程开放	687	424	1111
本科	1724	1974	3698
专科	2748	1812	4560
总计	5159	4210	9369

表二 各年龄段学生分布

年龄段	本科	课程开放	专科	总计
15~19	1	950	528	1479
20~24	1350	120	1730	3200
25~29	1035	15	717	1767
30~34	686	12	636	1334
35~39	369	7	397	773
40~44	193	7	341	541
45~49	56	—	165	221
50~54	8	—	38	46
55~59	—	—	7	7
不可知年龄	—	—	1	1
总计	3698	1111	4560	9369

学生登录行为描述与影响因素分析

通过对登录行为数据表做各种查询操作,可得到学生登录的基本情况信息,如表三所示。

表三 登录基本情况

学校名称	学生总数	平均登录次数	最大登录次数	平均登录天数	最大登录天数
江苏广播电视大学	9369	7	169	3天	56天

在登录行为数据表中,学生学号和登录时间是两个主要字段。下面笔者分别以登录时间(分“周”、“天”、“小时”等统计单位)作为分类字段,统计各个时段学生的登录人次;以学生学号为分类字段,统计每个学生的登录天数。本文不以学生的登录次数为计数单位,而是以某个时间单位的登录人数来统计。这是因为如果以登录次数来统计可能会由于个别学生的反复登录而造出学生学习强度大的假象。而登录人数能真实反映某天的学生整体活动程度。这里引入一个术语即登录率,作者将登录率定义为“登录率等于某一时间登录平台的学生数除以学生总数”。

(一) 以登录时间为分类字段的统计分析

1. 以“周”为单位统计每天的登录人数

以“2010-3-17”(课程开学日)所在周为第一周,接下来各周的起止结束时间如表四所示,一共有15周。笔者按周计算登录率,即“一周内学生登录的人数除以学生总数”,并

绘制“学习时间-周登录率”直方图,如图1所示。由图1可知,自开学后第4周起才陆续有学生登录网络课程进行学习,随即在第5周、第6周达到学习参与的高峰。自第9周起,学习登录率开始下降。除了学生登录率较高的4个周外(如第5周、第6周、第7周和第9周)以及极低的3个周(第1周至第3周),其他各周的登录率在10%至20%之间,相差不大。可以说,在整个学期,学生群体的学习是比较分散的。

表四 课程学习周期的各周及起止时间

相对周序	周开始时间	周结束时间
1	2010-3-14	2010-3-20
2	2010-3-21	2010-3-27
3	2010-3-28	2010-4-3
4	2010-4-4	2010-4-10
5	2010-4-11	2010-4-17
6	2010-4-18	2010-4-24
7	2010-4-25	2010-5-1
8	2010-5-2	2010-5-8
9	2010-5-9	2010-5-15
10	2010-5-16	2010-5-22
11	2010-5-23	2010-5-29
12	2010-5-30	2010-6-5
13	2010-6-6	2010-6-12
14	2010-6-13	2010-6-19
15	2010-6-20	2010-6-26

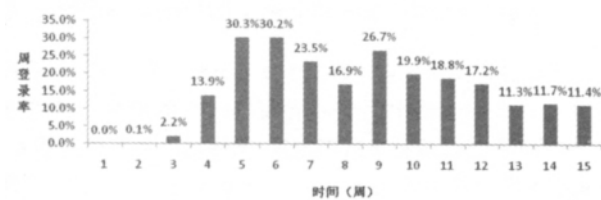


图1 各周学生登录率

2. 以“周”为单位统计发现群体学习的周期

对于学生个体而言,其学习周期是不确定的。但是从教学者的角度,有必要了解整个学生群体的学习周期,即通过登录平台的记录,查明多长时间内整个学生群体都会登录参与学习,这可以用来衡量群体在线学习开展的速度。另外,教学者还需要了解,随着学习时间的推移,有多少学生不再来学习了,这可以用来衡量群体在线学习终止的速度。通过对每周累计登录率的统计绘制学生的学习周期,得到如图2所示结果。由图2可知,直到开学第6周,才有50%的学生参与了学习。直到最后一周,所有学生才完成了一遍学习。这反映了学生整体开展在线学习的速度是很慢的。这就提

醒教学者必须加以干预,如在第2周就应该督促学生登录平台参加学习。

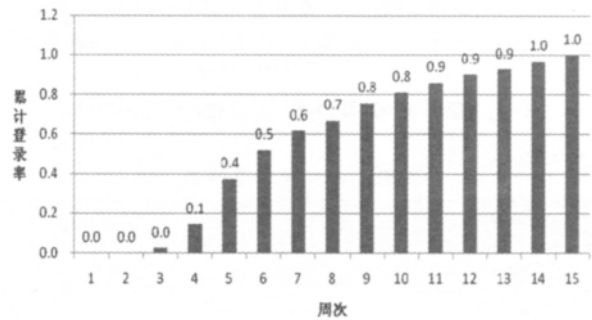


图2 学生累计登录率的时间分布

下面通过对每周累计流失的学生率的统计绘制学生的流失曲线,得到如图3所示结果。如图3所示,到11周,已有52%的学生不再来学习了。这一发现对于在不同时间合理分配教学资源有一定的指导意义。

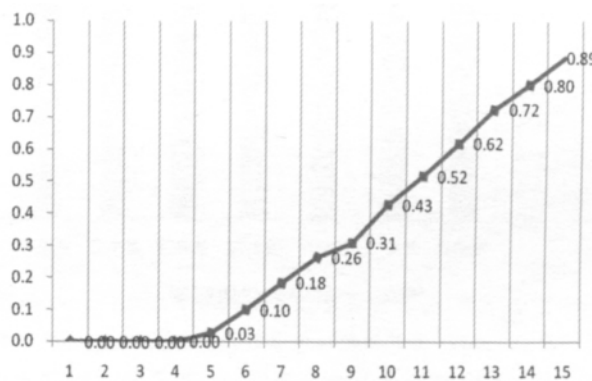


图3 学生流失曲线

3. 以“天”为单位统计每天的登录人数

前面以周为统计单位发现了群体学习的分散性,这种分散性还会表现在以更小的时间单位作为分类统计对象上。接下来以天为单位进行时间分布特点考察。研究者通过计算学生每天的登录率,绘制了“学习时间-登录率”直方图,得到如图4所示结果。同时列出单日登录率较高的几天的数据,如表五所示。由图4可知,在这个学习时段中,如果将学生日登录率不低于5%作为全体学生积极参与学习的标准,那么仅有26天时间里学生表现活跃。整个学习时段有103天,除去30天休息时间(15周),仍然有73天的学习时间。因此,学生整体活跃的时间只占整个学习时段的1/3。不过从另外一方面看,就群体角度而言,学生的在线学习时间是

比较分散的。笔者再以一周中的七天为分类统计项,计算学生登录率,得到如图5所示结果。如图5所示,学生分散在一周中的不同天学习。周六、周日的学习人数与工作日相当,并且在周日参与学习的学生比率最高,这体现了在职成人学习的特点,即利用节假日开展学习。

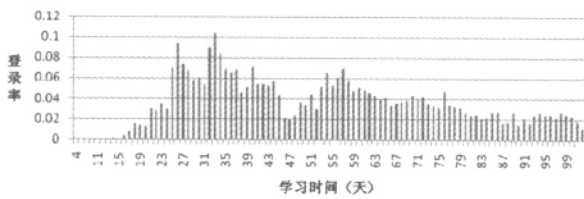


图4 “学习时间-登录率”直方图(以天为单位)

表五 单日登录率排名

日期	日登录人数	日登录率
2010-4-18	979	10.4%
2010-4-11	883	9.3%
2010-4-17	842	8.9%
2010-4-19	788	8.3%

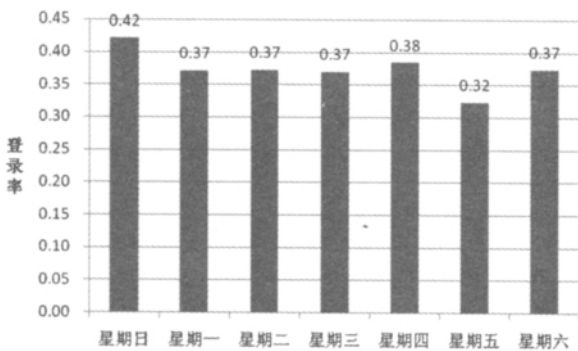


图5 一周中的登录时间

4. 以“时段”为单位统计每个时段的登录人数

接着我们再以时段(分为凌晨 00:00 - 6:00、上午 6:00 - 12:00、下午 12:00 - 18:00、晚上 18:00 - 00:00 等四个时段)为统计项,计算这四个时段的登录率,得到如表六所示结果。由表六可知,学生的学习时段主要有上午、下午和晚上三个时段,并且这三个时段的登录率相当,下午这一时段的登录率略高一些。

表六 学习时段分布

时段	登录学生数	登录率
凌晨	412	0.043975
上午	5627	0.600598
下午	6628	0.707439
晚上	5592	0.596862

(二) 以学生学号为分类字段的统计分析

前文以各种时间单位为分类字段统计得出在线学习中群体学习时间分散的特点。下面再以学生个体为分类字段进行统计,将会发现个体学习时间投入过少的特点。对于每个学习个体来说,都应该保证一定的学习时间,这样才能完成必要的学习任务。“开放教育学习指南”作为 1 学分的课程,其规定的学时是 18 学时,如果按照线上考核和线下考核

以 7: 3 的比例分配的话,笔者暂将线上学习和线下学习的学时也按照 7: 3 的比例进行分配,那么线上学习的学时应该是 12.6 学时。假设每个学生某天登录平台,其学习时间平均为 2 个学时的话,那么一个学生应该至少有 6 天登录平台参加学习。但事实并非如此。笔者首先查询得到每个学生有登录记录的天数,然后计算每类天数对应的学生数,最后建立累计学生数与登录天数的函数,如公式(1)所示。

$$R_s(t_i) = \sum^n S(t_i) / S_0 \quad \text{公式(1)}$$

其中 S_0 表示学生总数; t 表示登录天数,笔者将所有学生的登录天数按从高到底顺序排列并从 1 开始编号, t_i 即第 i 个登录天数; $S(t_i)$ 表示第 i 个登录天数对应的学生数; $R_s(t_i)$ 则表示截止到第 i 个登录天数,累计学生数占学生总数的比例。依照此公式以及实际数据,作者绘制得到了如图 6 所示的曲线。由图 6 可知,登录天数不少于 6 天的学生数不及学生总数的 20% (有 1420 名学生,占学生总数的 15.2%)。只有约 3% 的学生(258 名)的学习天数(指有登录记录的天数)不低于整个学习时间的 10% (11 天),而有近 80% 的学生(78.6%)的学习天数在 4 天及以下,50% 的学生的学习天数在 2 天及以下。这就反映出,大部分学生(超过 80%)未能投入必要的学习时间开展学习。

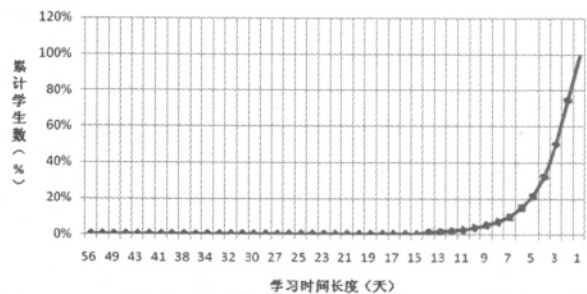


图6 登录天数与累计学生数的关系曲线

笔者更为关注登录天数仅为 1~2 天的学生群体(共计 4619 人),造成这些学生学习时间投入少的原因可能有两个,一是学生已经掌握了所需内容,无需投入过多时间学习;二是学生没有认真学习。我们有必要从“专业层次”、“性别”、“年龄”等学生属性进一步了解这些学生的特点。所采用的方法是 Microsoft 关联规则算法,以发现这 4619 位学生中三类属性组合的频繁项。笔者采用 SSAS(SQL Server Analysis Services) 部署挖掘结构和挖掘模型,将支持度设为 0.05,即每一属性组合对应的学生数不少于总人数的 5%,得到的 3-项集,如表七所示。这一人群存在于各个专业层次,其中男性略多于女性,以 29 岁以下年轻人为主。

另外,从整个学生群体来看,随着登录天数的减少,对应的学生数成倍增长。笔者发现,在一些登录天数范围内(如 1 天至 16 天)不同登录天数对应的学生数之间存在倍数关系,即登录天数每减少一天(或两天),减少之后的天数对应的学生数就会增加 1 倍,用公式可以简单表示为:

表七 登录天数少的学生群体主要特征

项集维度	项集	学生数
3	专业层次 = 专科, 年龄值 < 22, 性别 = 男	502
3	专业层次 = 专科, 年龄值 = 22-29, 性别 = 男	499
3	专业层次 = 本科, 年龄值 = 22-29, 性别 = 女	469
3	专业层次 = 本科, 年龄值 = 22-29, 性别 = 男	455
3	专业层次 = 课程开放, 年龄值 < 22, 性别 = 男	405
3	专业层次 = 专科, 年龄值 < 22, 性别 = 女	307
3	专业层次 = 课程开放, 年龄值 < 22, 性别 = 女	281
3	专业层次 = 专科, 年龄值 = 29-36, 性别 = 男	247
合计		3165

$$S(t_{i-1}) = S(t_i) * 2 + a \quad \text{公式(2)}$$

这一公式可通过如表八所示的数据加以验证。如表八所示, 笔者将登录天数两两进行合并, 如“1天”和“2天”合并为“1.5”天, “3天”和“4天”合并为“3.5天”, 以此类推, 然后通过“ $S(t_{i+1}) / S(t_i) * 2$ ”的计算来得到倍数关系。可以发现, 这些倍数在2左右摆动。这种现象类似于无尺度网络现象。张超等(2009)曾对两个网络学习课程班的异步交互网络进行比较分析, 结果发现学习者接近自然条件下(即缺少教师干预)的在线交互行为构成无尺度网络, 而强学习干预手段的实施使得学习者交互行为发生改变。由此, 我们可以推测, 产生登录天数对应的学生数呈偏态分布的原因可能是由于该网络课程的开展缺少教师的干预, 学生在放任状态下学习。

表八 登录天数对应的学生数

序号	登录天数	登录天数均值	对应学生数总和	倍数关系
1	16	15.5	35	
	15			
2	14	13.5	53	1.514286
	13			
3	12	11.5	112	2.113208
	11			
4	10	9.5	253	2.258929
	9			
5	8	7.5	447	1.766798
	7			
6	6	5.5	1049	2.346756
	5			
7	4	3.5	2698	2.571973
	3			
8	2	1.5	4619	1.712009
	1			

(三) 影响学生登录天数的因素

接下来笔者拟采用相关分析、差异检验、回归分析等方法, 以发现影响学习者学习时间投入的相关因素。分析工具采用 SPSS13.0。这里的因变量是学生的登录天数(即某个学生有登录记录的天数), 自变量则是学生的基本属性, 包括年龄、性别和专业层次。

1. 登录天数与年龄段之间的关系

对表二所划分出的各年龄段学生对应的登录天数做描述性统计, 得到如表九所示结果。

表九 各年龄段学生登录天数的描述性统计

年龄段	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
15 ~ 19	1479	2.9932	3.50738	.09120	1.00	44.00
20 ~ 24	3200	3.2413	3.33630	.05898	1.00	52.00
25 ~ 29	1767	3.5750	3.24236	.07713	1.00	40.00
30 ~ 34	1334	3.8111	3.52532	.09652	1.00	56.00
35 ~ 39	773	3.8629	3.53580	.12717	1.00	41.00
40 ~ 44	541	4.1442	3.75715	.16153	1.00	40.00
45 ~ 49	221	4.1312	3.41999	.23005	1.00	18.00
50 ~ 54	46	3.5217	2.87316	.42362	1.00	15.00
55 ~ 59	7	2.7143	1.11270	.42056	1.00	4.00
合计	9368	3.4716	3.43141	.03545	1.00	56.00

接着, 笔者对各年龄段学生的登录天数做差异检验, 采用 One-way ANOVA 单因子变异数分析方法, 得到如表十所示结果。由表十可知, 年龄较小的学生(低于25岁)与年龄较大的学生(不低于25岁)的登录天数均值存在显著差异, 年龄较大的学生(不低于25岁)的平均登录天数多于年龄较小的学生(低于25岁)的平均登录天数。

表十 各年龄段之间的差异

平均数差值 年龄段	年龄段						
	15 ~ 19	20 ~ 24	25 ~ 29	30 ~ 34	35 ~ 39	40 ~ 44	45 ~ 49
15 ~ 19		0.25	0.58 (*)	0.82 (*)	0.87 (*)	1.15 (*)	1.14 (*)
20 ~ 24	0.25		0.33 (*)	0.57 (*)	0.62 (*)	0.90 (*)	0.89 (*)
25 ~ 29	0.58 (*)	0.33 (*)		0.24	0.29	0.57	0.56
30 ~ 34	0.82 (*)	0.57 (*)	0.24		0.05	0.33	0.32
35 ~ 39	0.87 (*)	0.62 (*)	0.29	0.05		0.28	0.27
40 ~ 44	1.15 (*)	0.90 (*)	0.57	0.33	0.28		0.01
45 ~ 49	1.14 (*)	0.89 (*)	0.56	0.32	0.27	0.01	

* The mean difference is significant at the .05 level.

各年龄段与对应的登录天数均值的相关分析结果表明, 二者存在明显的正相关, 如表十一所示。这里为排除干扰, 在计算相关性时剔除了学生样本较少的年龄段, 如50~54

和 55~59 两个年龄段。表十一所示的计算结果表明,随着年龄段的增长,各年龄段的学生的学习投入时间在不断增加。

表十一 相关分析

	登录天数均值	年龄段
登录天数均值	1	.973(**)
年龄段	.973(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

2. 登录天数与性别之间的关系

对不同性别学生的登录天数的差异检验结果显示,不同性别学生的登录天数存在显著差异。如图 7 所示,独立样本 t 检验结果显示,不同性别的学生在“登录天数”方面存在显著差异(由于 F 值达到显著, t 值取“Equal variances not assumed”列之 t 值,此时 t = -6.047, Sig. < 0.05),女生的平均登录天数多于男生的平均登录天数。

表十二 不同性别学生的登录天数统计

性别	登录天数	样本数	均值	标准差	标准误
男		5159	3.2789	3.45810	.04815
女		4210	3.7081	3.38355	.05215

Legend	Levene's Test for Equality of Variances		t Test for Equality of Means						
	f	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Equal variances assumed	14.086	.000	-6.033	9367	.000	-.42915	.07113	-.58658	-.28971
Equal variances not assumed			-6.047	9066.913	.000	-.42915	.07067	-.58627	-.29002

图 7 独立样本 t 检验结果

3. 登录天数与专业层次之间的关系

通过对学生的登录天数按照专业层次(本科、专科和课程开放三类)分组做差异分析,采用 one-way ANOVA 单因子变异数分析方法,结果发现不同专业层次之间的登录天数存在显著差异,如表十三和表十四所示。如表十四所示,由单因子变异分析结果可知,“课程开放”学生与“本科”学生在“登录天数”方面存在显著差异,“课程开放”学生与“专科”学生在“登录天数”方面存在显著差异,“本科”学生与“专科”学生在“登录天数”方面不存在显著差异。本科、专科学生的平均登录天数多于“课程开放”学生的平均登录天数。

表十三 不同专业层次学生的登录天数统计

	N	Mean	Std. Deviation	Std. Error	Minimum	Maximum
本科	3698	3.6022	3.11765	.05127	1.00	56.00
专科	4560	3.5605	3.86449	.05723	1.00	52.00
课程开放	1111	2.6733	2.21957	.06659	1.00	32.00
总计	9369	3.4718	3.43126	.03545	1.00	56.00

学生资源浏览行为描述与影响因素分析

“开放教育学习指南”网络课程包含课程各章节、体验

区、视频资源库、问题库、资料库和常用工具库等内容模块,这些模块是学生开展学习的主要对象,其内容分布在若干网页上。每个网页均有记录打开页面和关闭页面的功能,最终形成页面浏览日志。笔者依据页面浏览日志统计分析该课程各模块的浏览情况和不同模块中各页面的浏览情况,以发现各类资源的使用情况并分析影响资源使用的相关因素。如表十五所示,是网络课程各模块的一个说明。

表十四 单因子变异数分析结果

(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
本科	专科	.04169	.07683	.930	-.1418	.2252
	课程开放	-.92895 (*)	.08404	.000	-.7282	1.1297
专科	本科	-.04169	.07683	.930	-.2252	.1418
	课程开放	.88726 (*)	.08780	.000	.6775	1.0970
课程开放	本科	-.92895 (*)	.08404	.000	-1.1297	-.7282
	专科	-.88726 (*)	.08780	.000	-1.0970	-.6775

表十五 “开放教育学习指南”课程各模块说明

模块名称	模块说明
课程各章节	网络课程主体模块,包含有 61 个章节内容页面
体验区	模拟网络平台使用过程,有 7 个过程模拟页面
视频资源库	列出了课程各章节模块出现的视频资源
问题库	列出了网络学习中存在的常见问题
资料库	列出了开放教育的一些教学管理文件
常用工具	列出了网络学习中可能用到的一些工具

(一) 以课程各模块为分类字段的统计分析

表十六是“开放教育学习指南”网络课程各模块的浏览情况。

• 被浏览的资源页面个数:通过页面浏览日志数据统计得到的各模块页面个数,可能少于各模块实际包含的页面数。

• 单个页面浏览频次均值:由“浏览频次/被浏览的资源页面个数”算式得来,反映该模块各资源页面的使用程度。

• 浏览学生数:通过页面浏览日志数据统计得到的浏览各模块的学生数。

• 学生参与率:由“浏览学生数/学生总数”算式得来,用来反映该模块被学生关注的程度。

• 人均浏览频次:由“浏览频次/浏览学生数”算式得来,用来反映学生在该模块的学习程度。

• 单个资源人均浏览频次:由“人均浏览频次/被浏览的资源页面个数”算式得来,用来反映学生在该模块各资源页面的学习程度。

表十六 网络课程各模块浏览情况

模块名称	被浏览的资源页面个数	浏览频次	单个页面浏览频次均值	浏览学生数	学生参与率	人均浏览频次	单个资源人均浏览频次
课程各章节	103	809797	7862.11	9230	0.99	87.74	0.85
体验区	9	206720	22968.89	9049	0.97	22.84	2.54
视频资源库	37	36978	999.41	4321	0.46	8.56	0.23
问题库	6	34889	5814.83	3262	0.35	10.70	1.78
资料库	49	15843	323.33	2761	0.29	5.74	0.12
常用工具库	11	634	57.64	382	0.04	1.66	0.15

如表十六所示,从模块的被使用程度和被学生关注的程度来看,“浏览频次”列和“学生参与率”列,由高到底依次是课程各章节、体验区、视频资源库、问题库、资料库、常用工具库,这恰好是这些模块在课程首页由上而下呈现的顺序。由此可见,模块在页面中的布局对于模块的被使用程度有重要影响。对于网络课程界面设计的意义则在于,要想使某个模块得到学生关注,应该将其置于首页的显著位置。

从学生对不同模块各个资源页面的学习程度来看(观察“单个资源生均浏览频次”),由高到底依次是体验区、课程各章节、问题库、视频资源库、资料库、常用工具库。学生对体验区、课程各章节各资源页面的学习程度高应该是受到了作业1和作业3的导向作用影响。相比之下,体验区的学习程度更高一些,每个学生对体验区的每个页面的浏览次数平均为2.54次,而课程各章节的单个资源生均浏览频次则在1次以下。产生这一差异的原因应该是:体验区采用一种模拟学习方式,以互动的方式让学生掌握网上学习技能,应该更能引发学生的学习兴趣,吸引学生反复学习。对于其他四个非作业要求的模块来说,学生对问题库的学习程度明显高于其他三个模块,说明学生有较强的通过常见问题库来解答初入学各种疑问的需求。今后应该进一步丰富问题库,并将问题库的检索框置于首页的显著位置,以便及时、方便地解答学生的各种疑惑,以帮助学生尽快进入开放教育之门。

视频资源库模块被学生关注的程度不高(学生参与率为0.46,即不到一半的学生浏览了该模块),视频资源库模块中的各个视频资源的被学习程度也很低(单个资源人均浏览频次为0.23)。由于视频资源模块中的各个视频资源都是课程各章节内容页面的一部分,如果学生认真花时间学习课程各章节页面的话,应该会观看页面上的视频资源。但是浏览视频资源模块的学生比例只占浏览课程各章节模块的学生比例的一半不到,而具体到模块中的各个页面,视频资源模块的学习程度也要显著低于课程各章节模块的学习程度。这只能说明学生对课程各章节的学习并不彻底,很多资源特别是视频资源都没有看。龚志武(2004)通过对学生的调查发现,流媒体课件(即本文中的视频资源)不能单独发挥作用,它必须紧密地与课程内容相结合,与导学指引相结合,与网络教学环境相结合,才能真正发挥作用,否则就成为摆设。因此,“开放教育学习指南”网络课程的视频资源未能得到充

分应用,其原因很可能在于这些视频资源未能很好的与课程内容相结合,课程设计还需加以改进。

(二) 以学生学号为分类字段的浏览行为统计分析

该课程有三个作业,这三个作业通过浏览模块“课程各章节”和“体验区”来完成。据统计,这两个模块有实际内容(目录页除外)的资源页面分别为61个和7个。另外,视频资源库中的28个视频实际上是“课程各章节”模块的重要组成部分,应该成为在作业1中完成的内容。综上,形成性评价所要求浏览的、有实际内容的资源页面共计96个,笔者将这些资源定义为“有效资源”。下面就以各个学生为统计项,调查他们对这96个有效资源的浏览情况,以发现学生在使用资源程度的分布情况以及时间分散特点。表十七是学生浏览有效资源的基本情况。

表十七 学生浏览有效资源基本情况

学生参与数	9231人	登录天数(累计)	32527天
学生参与率	0.985	人均浏览资源数	68个
未浏览资源人数	138人	浏览资源数众数	68个
人均浏览资源天数	2天	最大浏览资源数	96个
最大浏览资源天数	35天	人均浏览资源频次	84次
浏览资源天数(累计)	18852天	最大浏览资源频次	970次

由表十七可知,在所有有登录记录的学生中,仍然有138人未浏览任何有效资源,即这些学生的在线学习没有发生。从人均浏览资源天数和最大浏览资源天数来看,二者均小于与之相应的平均登录天数和最大登录天数。另外,比较学生浏览有效资源的天数(累计为18852天)和学生登录的天数(累计为32527天),前者为后者的58%,这说明学生在有近一半的“学习日”(学生进入平台的日子)只是路过平台而已,并没有发生有效的学习,也表明学生在开展在线学习过程中有缺少计划、随意中止的特点。从登录天数来看,学生有意来学习并登录平台的时间是很有限的,如平均登录天数为3天,近80%的学生(78.6%)的学习天数在4天及以下。在这有限的登录天数里,又有一半的时间放弃了学习,这是非常可惜的。如果不能抓住每一次登录时间及时学习,势必会导致最后的突击学习,即在1~2天内浏览完所有有效资源。这一突击学习的特点将在后面加以验证。

从浏览资源数的情况来看,最大浏览资源数是96个,说明有的学生浏览了所有资源。人均浏览资源数为68个,并且“68个”也是所有学生浏览资源数中出现次数最多的,结合前面对模块有效资源数的统计,这“68个”对应的资源主要来自“课程各章节”和“体验区”两个模块,而这两个模块的有效资源数之和恰好是68个。这说明大部分学生在开展在线学习时以完成作业为最高要求。

下面我们再来看不同资源浏览数的学生分布情况,以了解整个学生群体中对有效资源的浏览情况,如表十八所示。可以看出,以完成作业1和作业2为最高要求的学生数量最多,即资源浏览数在68个左右(对应的范围为60~79)这部分学生占全部学生(9369名)的83%。而浏览了60个以下资源的学生约为7.2%,只是很少的一部分学生。看来,绝大部分学生能够完成作业。但这些完成了作业的学生,他们是以何种方式来浏览这些资源的,是集中在一两天突击完成的,还是在数天之中分散浏览的?接下来,笔者对资源浏览数在60个及以上的学生对应的浏览天数进行统计,得到如表十九所示结果。如表十九所示,只有不及1/4的学生有3天及以上的浏览天数。而超过3/4的学生在1~2天时间里完成了几乎所有有效资源的浏览,足见突击学习的普遍性。

表十八 学生资源浏览数分布情况

资源浏览数	学生数	学生比例	累计学生比例
90~99	266	0.03	2.9%
80~89	671	0.07	10.0%
70~79	2112	0.23	32.6%
60~69	5646	0.60	92.8%
50~59	114	0.01	94.1%
40~49	89	0.01	95.0%
30~39	48	0.01	95.5%
20~29	56	0.01	96.1%
10~19	72	0.01	96.9%
0~9	157	0.02	98.6%

最后,通过对每个学生的登录天数与资源浏览天数做Person相关分析,相关系数为0.661($\text{sig} < 0.01$),属显著正相关。因此,学生群体中登录天数的分布特点与资源浏览天数的分布特点存在相似性。如果从学生基本信息的角度发现资源浏览天数的影响因素,所得结果应该与前面对登录天数影响因素的分析是相似的。所以,这里不再做重复分析。

(三) 从模块访问跳转的角度看浏览行为

这里采用Microsoft顺序分析和聚类分析算法,从模块访问跳转的角度分析学生的学习路径。数据来源则是学生每天浏览课程页面产生的过程数据,如表二十所示。

在构建挖掘模型时,以“浏览日期+学号”作为键值,以“浏览顺序号”作为序列键值,以“浏览模块”作为预测值来构建挖掘结构,应用Microsoft顺序分析和聚类分析算法,得

表十九 登录天数及与之对应学生数的分布

浏览天数	学生数	累计学生比例
32	1	0.01%
21	1	0.02%
19	2	0.05%
18	4	0.09%
17	2	0.12%
16	2	0.14%
15	2	0.16%
14	1	0.17%
13	2	0.20%
12	10	0.31%
11	8	0.40%
10	17	0.60%
9	28	0.92%
8	31	1.28%
7	67	2.05%
6	112	3.34%
5	238	6.07%
4	510	11.94%
3	1067	24.21%
2	2618	54.32%
1	3972	100.00%

表二十 某学生的课程浏览样例

浏览日期	浏览页面所属模块	浏览顺序号
04/27/2010	课程章节	1
04/27/2010	课程章节	2
04/27/2010	体验区	3
04/27/2010	体验区	4
04/27/2010	体验区	5

到如表二十一和图8所示结果。从中可以看出,学生在登录网络课程后,最有可能先浏览的是“课程章节”模块,其次是“体验区”模块,极少从其他模块开始浏览。图8是各模块间跳转的状态转换图。可以看出,学生一旦进入“课程章节”或“体验区”模块,主要是在本模块活动,期间几乎不去访问其他模块。“问题库”、“资料库”、“常用工具库”等学习辅助模块在课程主体部分即“课程章节”学习过程中几乎没有被用到(学生偶尔从这三个模块进入,然后转入“课程章节”模块,而不是反过来)。这可以从课程的评价导向和“课程章节”部分的链接设计来找原因。从评价导向来看,教师只要

求浏览“课程章节”和“体验区”模块,其他模块就被忽略了。而“视频资源库”由于与“课程章节”紧密结合,且在“课程章节”有链接,因而得到了较好的利用,从状态转换图来看两个模块来回访问的概率不小。这就反映出其他三个辅助模块得不到访问可能存在内容设计上的问题,即这三个模块与“课程章节”主模块之间相互孤立,没有链接。

表二十一 各模块(PAGETYPE)之间的转换概率

变量	值	概率
PAGETYPE. 转换	[开始]-> 课程章节	68.525%
PAGETYPE. 转换	课程章节,课程章节	67.734%
PAGETYPE. 转换	[开始]-> 体验区	29.346%
PAGETYPE. 转换	体验区,体验区	16.624%
PAGETYPE. 转换	课程章节,视频资源库	2.406%
PAGETYPE. 转换	视频资源库,课程章节	2.241%
PAGETYPE. 转换	问题库,问题库	1.988%
PAGETYPE. 转换	资料库,资料库	0.845%
PAGETYPE. 转换	体验区,课程章节	0.832%
PAGETYPE. 转换	视频资源库,视频资源库	0.822%
PAGETYPE. 转换	[开始]-> 问题库	0.806%
PAGETYPE. 转换	问题库,课程章节	0.731%
PAGETYPE. 转换	[开始]-> 资料库	0.632%
PAGETYPE. 转换	缺失	0.631%
PAGETYPE. 转换	[开始]-> 视频资源库	0.532%

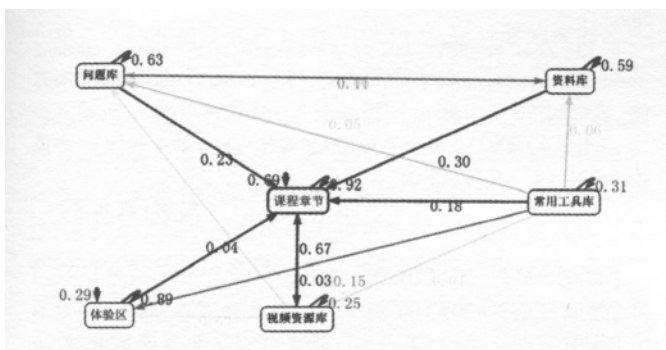


图8 各模块的跳转情况

形成性考试成绩分析

表二十二是对考试成绩数据的一个描述性统计。总体而言,学生的成绩是很高的,各项作业及总分的平均成绩几乎都达到或超过了满分的90%,这与前面所得出的学生行为分析结果是不一致的,这反映出该网络课程考核方式的效度有待进一步改进。由于形成性考试成绩从学生行为记录转换而来,前面已对学生行为表现的影响因素做了分析,这里不再做重复分析。

表二十二 三项作业及总成绩的描述性统计

统计项	N	满分	Minimum	Maximum	Mean	Std. Deviation
总分	9369	70.00	.00	70.00	63.3370	9.75347
作业一	9369	15.00	.00	15.00	14.4227	2.51660
作业二	9369	40.00	.00	40.00	35.2695	6.39015
作业三	9369	15.00	.00	15.00	13.6448	3.80213

研究结论与建议

本文以“开放教育学习指南”网络课程为例,以2010年春季学期江苏电大9369名新生学习该课程时所产生的登录数据、资源浏览数据和形成性测试数据为分析对象,发现了登录、资源浏览、作业、测试等在线学习行为的基本特点及其影响因素。研究结果表明:

1) 从群体角度看,学生开展在线学习的时间比较分散。学生的学习活动较为均匀地分散在一天中的三个时段(下午时段学生登录率最高)、一周中的七天(周日学生登录率最高)、一学期中的15周(第5周、第6周学生登录率最高),没有学生登录率特别集中的某一周或某一天(单周最高登录率为30.3%,单日最高登录率为10.4%)。

2) 从个体角度看,学生在线学习时间投入低,其外在因素之一可能是缺乏必要的教学干预,近80%的学生(78.6%)登录天数在4天及以下,50%的学生的登录天数在2天及以下。其中,倾向于“登录天数”“不少于3天”的学生有“专业层次=本科”、“专业层次=专科”、性别=女、年龄=不小于25岁”和“专业层次=课程开放,年龄=不小于25岁”等三个群体;倾向于“登录天数”“少于3天”的学生有“专业层次=专科,性别=男,年龄=小于25岁”和“专业层次=课程开放,年龄=不小于25岁”等两个群体。

3) 学生浏览资源以完成作业为最高要求,并且在1~2天内突击完成所有资源的浏览任务;在线学习呈现出缺少计划、随意终止的特点,有一半的登录时间学生虽然进入了网络课程平台却未浏览有效资源。83%的学生的资源浏览数刚刚达到形成性考核的要求,而在完成了形成性考核所要求浏览的资源数的学生当中,有超过3/4的学生在1~2天内完成了几乎所有有效资源的浏览,反映出学生的在线学习集中性、突击性的特点。

4) 学生一旦进入“课程章节”或“体验区”等课程主要模块,主要是在本模块活动期间几乎不去访问其他模块。“问题库”、“资料库”、“常用工具库”等学习辅助模块在课程主体部分即“课程章节”学习过程中几乎没有被用到(学生偶尔从这三个模块进入,然后转入“课程章节”模块,而不是反过来)。这可以从课程的评价导向和“课程章节”部分的链接设计来找原因。从评价导向来看,教师只要求学生浏览“课程章节”和“体验区”模块,其他模块就被忽略了。而“视频资源库”由于与“课程章节”在内容上紧密结合,且在“课

程章节”中有链接,结果得到了较好利用,从状态转换图来看,两个模块来回访问的概率不小。这就反映出其他三个辅助模块得不到访问可能存在内容设计上的问题,即这三个模块与“课程章节”模块之间相互孤立,没有链接。

5) 学生三次形成性作业得分及最后的总分普遍很高,平均分接近或超过满分的占90%,这与学生在登录和资源浏览上的表现是不一致的,表明该课程形成性评价方案的效度有待进一步验证。

6) 影响学生学习时间投入的内在因素有性别、年龄和专业层次,其中女生比男生投入的学习时间多,年长学生比年轻学生投入的学习时间多,本科学生比专科学生、专科学生比课程开放学生投入的学习时间多。

上述研究结论给我们以下启示:中央电大开放教育学生在线学习时间投入偏低,需要辅导教师提出严格的教学管理制度,以强制性手段要求学生加大在线学习投入,只有投入时间才可能引发学习;网络课程的设计有待改进,应增强网络课程各个模块的超链接,可将辅助模块的内容无缝嵌入课程主模块中,提高课程辅助模块的利用率;在线学习评价效度有待进一步优化,要使评价既能促进学生的学习,又能测量出学生的学习结果。

【参考文献】

- [1]Gong Zhiwu(2004). An empirical study on impacting factors of adult learner's online learning behavior(in Chinese) [J]. China Educational Technology, (8): 32-35.
(龚志武(2004). 关于成人学生网上学习行为影响因素的实证研究[J]. 中国电化教育, (8): 32-35.)
- [2]Ha S., Bae S., & Park S. (2000). Web mining for distance education[A]. IEEE international conference on management of innovation and technology[C]. 715-719.
- [3]Mao Lijie & Cao Xiaoying(2009). Research and design of monitoring learning function based on online learning platform(in Chinese) [J]. Software Guide, (9): 79-81.
(毛利杰,曹效英(2009). 基于网络教学平台的学习监控功能研究与设计[J]. 软件导刊(教育技术), (9): 79-81.)
- [4]Romero, C., Ventura, S., & Garcia, E. (2005). Data mining in course management systems: Moodle case study and tutorial[J]. Computers & Education 51(1): 368-384.
- [5]Shao Yanli, & Sun Lina(2008). Monitoring learners behavior in e-portfolio-based online learning(in Chinese) [J]. China Distance Education, (5): 38-40, 80.
(邵艳丽,孙丽娜(2008). 基于电子学档的在线学习行为监控[J]. 中国远程教育, (5): 38-40, 80.)

[6]Song Jiangchun & Chen Wenlin(2005). Web data mining and it's application in learning support service of distance education(in Chinese) [J]. China Distance Education, (7): 62-64.

(宋江春,陈文林(2005). Web使用挖掘及其在远程教育教学支持服务中的应用研究[J]. 中国远程教育, (7): 62-64.)

[7]Wang F. (2002). On using data-mining technology for browsing log file analysis in asynchronous learning environment[A]. Conference on education multimedia, hypermedia and telecommunications [C]. 2005-2006.

[8]Xiao Aiping, & Jiang Chengfeng(2009). Investigation and research on web-based learners(in Chinese) [J]. Open Education Research, 15(1): 75-80.

(肖爱平,蒋成凤(2009). 网络学习者网上学习现状、影响因素及对策研究[J]. 开放教育研究, 15(1): 75-80.)

[9]Xie Weiqi(2004). Web data mining based on distance education platform(in Chinese) [J]. China Edu Info, (10): 62-56.

(谢维奇(2004). 基于“电大在线”远程教学平台的Web数据挖掘[J]. 教育信息化, (10): 62-56.)

[10]Yang Jinlai, Zhang Yixiang & Ding Rongtao(2008). A Study on Monitoring Learners' Learning Behavior based on Online Learning Platform(in Chinese) [J]. Computer Education, (11): 65-68.

(杨金来,张翼翔,丁荣涛(2008). 基于网络学习平台的学习行为监控研究[J]. 计算机教育, (11): 65-68.)

[11]Yuan Ming & Chen Weijie(2006). A study on analyzing on-line learning behavior of learner from online education(in Chinese) [J]. Zhejiang Educational Technology, (4): 6-9.

(袁明,陈伟杰(2006). 网络教育学习者在在线学习行为分析研究[J]. 浙江现代教育技术, (4): 6-9.)

[12]Zhang Chao & Zhu Zhiting(2009). Topology study on online learners' asynchronous interaction: An analysis based on complex network model(in Chinese) [J]. e-Education Research, (2): 59-63.

(张超,祝智庭(2009). 在线学习者异步交互的拓扑结构研究——一种基于复杂网络模型的分析[J]. 电化教育研究, (2): 59-63.)

[13]Zhao Peiyuan(2002). An analysis of the factors affecting students distance learning(in Chinese) [J]. China Distance Education, (8): 50-52, 79.

(赵丕元(2002). 影响学生远程学习行为因素的分析[J]. 中国远程教育, (8): 50-52, 79.)

(编辑:魏志慧)

【收稿日期】 2011-01-10

【修回日期】 2012-06-23

【作者简介】 魏顺平,博士,副研究员,中央广播电视大学现代远程教育研究所,研究方向为教育数据挖掘(wsp2049@163.com)。

(下转第17页)

