



学习分析视角下 MOOCs 形成性测试 难度系数研究*

宗 阳 郑勤华 张 玄 陈 丽

(北京师范大学 远程教育研究中心, 北京 100875)

[摘 要] 随着学习分析技术的发展,学界对 MOOCs 学习评价研究越来越侧重对其学习过程的评价研究,并且多采用形成性评价的方式。在传统在线测试题库中,难度适宜的测试题能帮助学习者超越其自身最近发展区,调动其学习积极性。在 MOOCs 中,学习者的最终学习成绩多由平时作业测试和期末测试成绩按照一定权重综合给定,对学习者的学习进行形成性评价主要依据形成性测试成绩。因此,在 MOOCs 中,形成性测试的难易程度如何?在学习分析视角下,怎样基于学习者形成性测试成绩,分析 MOOCs 形成性测试的难易程度?研究发现,利用统计学上计算传统测试题难度系数小样本收敛的方法,结合在 MOOCs 中形成性测试与传统测试题的相似性,可以构建 MOOCs 形成性测试难度系数自动学习收敛模型,该模型在一门中国 MOOC 中得到了科学性和可行性验证。MOOCs 形成性测试难度系数的研究方法,可以为 MOOCs 中进行形成性测试设置时提供一定的参考依据。

[关键词] 学习分析;MOOCs;形成性测试;难度系数

[中图分类号] G434 [文献标识码] A [文章编号] 1672-0008(2016)03-0096-08

DOI:10.15881/j.cnki.cn33-1304/g4.2016.03.011

一、引言

大规模开放式网络课程 (Massive Open Online Courses, MOOCs) 作为互联网与教育结合的产物,在最近几年得到了飞速发展。MOOCs 并非简单地将传统教育从线下搬到线上,而是一种新型的教育生态^[1]。在这种新型教育生态下,通过学习分析技术以及教育数据挖掘的方法进行教育网络信息生态系统的测评,可以发现系统构成要素以及要素间的协调程度,从而有针对性地提出系统优化和配置策略^[2]。尽管行为主义的教学模式近年来屡遭诟病,但其更加接近于传统教学模式,基于行为主义的 xMOOCs 在 MOOCs 中仍然占据主导地位^[3],因此,在本研究中的 MOOCs 均指基于行为主义的 xMOOCs。MOOCs 高辍学率的弊端,使得它的质量保证问题在近年来受到越来越多的教育研究者及社会人士关注,如何采用合理的形成性评价手段来保证高质量的教学,是以 MOOCs 为代表的在线教育面临的重大难题。

在美国新媒体联盟最新发布的《地平线报告》(2016 高教版)中,把学习分析和自适应学习作为当

前 1 年内的教育技术的主要发展趋势^[4]。学习分析是对学习者的学习行为进行分析,从而对学习者的学习进行描述、诊断、预测与干预。传统的教学质量一般通过学习者的课堂表现情况以及学习成绩进行反映;通常采用形成性测试和总结性测试两种方式,对 MOOCs 中的教学质量进行检测,总结性测试侧重于对学习者的学习结果进行评价,而科学合理的形成性测试可以有效检测学习者的阶段性学习过程,进而反映学习者学习的过程性。

随着学习分析技术的发展,对 MOOCs 学习者学习过程的评价日益受关注,并且其多采用形成性评价的方式。在传统测试中,合理设置符合学习者认知发展水平的测试题能激发学习者的学习积极性^[5],在 MOOCs 中的形成性测试作为形成性评价的主要方式是对学习者阶段性学习内容掌握情况的反馈^[6]。而合理设置形成性测试题对 MOOCs 学习者的学习有重要影响,难度适宜的测试题能帮助学习者跨越其自身的最近发展区,激发学习者的学习潜能,而成功进入下一个学习内容;对教师或管理者而言,难度

* 基金项目:本文系全国教育科学“十二五”规划 2014 年度国家重点课题“教育信息化与大型开放式网络课程(MOOCs)战略研究”(ACA140009)的研究成果。

适宜的形成性测试可以对学习者进行有效考核,进而结合教学目标改进教学,提高在线教学质量。

MOOCs 学习者的最终学习成绩多由平时单元作业测试和期末测试成绩按照一定比例综合评定,MOOCs 中的作业测试及期末测试是形成性测试的主要组成部分。MOOCs 形成性测试的难易程度如何?学习分析技术的迅速发展,使得对在线学习者学习过程中产生的学习行为的测量记录变得更加精确,在 MOOCs 中对学习者形成性测试成绩有较为详尽的记录。在学习分析视角下,怎样基于学习者形成性测试成绩分析 MOOCs 形成性测试的难易程度?本研究将 MOOCs 形成性测试中计入学习者最终成绩计算的每个单元作业测试,类比为传统测试中的每道测试题,采用统计领域相对成熟的难度系数计算方法,对国内 MOOCs 形成性测试难易程度进行判断,基于 MOOCs 学习者形成性测试成绩,结合失分率、区分度和信度形成 MOOCs 形成性测试难度系数计算模型,并将该模型应用于一门中国 MOOC 进行科学性和可行性验证。

二、相关研究

(一)学习分析与形成性评价研究

在首届“学习分析和知识国际会议”中,对学习分析进行了定义:学习分析技术是测量、收集、分析和报告有关学生的学习行为以及学习环境的数据,用以理解和优化学习及其产生的环境的技术。美国新媒体联盟发布的《地平线报告》认为,学习分析技术为推动高等教育领域的改革,促进个性化学习,开展形成性评价以及绩效测量等提供了有效的支持,提供了利用统计和数据深度学习活动的工具,从而提高学生的学习效果和个性化的学习体验^[7]。

学习分析技术在教育领域有着广阔的应用前景和发展潜力,它可以作为教师教学决策、优化教学的有效支持工具,也可以为学生的自我导向学习、学习危机预警和自我评估提供有效数据支持,还可以为教育研究者的个性化学习设计和增进研究效益提供数据参考^[8]。学习分析技术被广泛应用于对学习者的知识、行为和经历建模,学习者建档,领域知识建模以及趋势分析等研究领域^[9]。学习分析中的分析的数据包括结构化数据和非结构化数据,结构化数据往往指作业及测验成绩,在学习过程中产生的时间、次数等数据;非结构化数据一般指讨论区的帖子、作业

内容等量化的质性数据^[10]。

国内外众多学者应用学习分析技术,对学习者学习活动中产生的数据进行统计和深度分析研究,从而指导教学。比如,顾小清等人通过学习分析 Sakai 平台记录学生访问情况预测高校学生学业成就,用 iPad 记录小学生形成性“注意”表现,从而协助教师发现学生的学习困难^[11]。武法提等人通过对学习者及其在所在情景中产生的数据进行分析,构建了基于学习者个性行为分析的学习结果预测框架^[12]。在应用学习分析技术进行研究的同时,还产生了一批自动化的学习分析工具。比如,美国马兰里大学巴尔的摩分校的“Check My Activity”,通过 BB 平台对学习者在网络学习中的学习活动和适用频率进行检测,获取其学习过程中的行为数据,根据学习者在网络学习中的表现,对学习者的学习情况进行登记评价,从而帮助教师了解学生的学习情况^[13]。美国普渡大学的“Course Signals”系统通过对学习者在学习过程中的课程表现,课程努力程度,前期学业历史以及学习者特征这几个方面的行为数据进行分析从而对学习者的学习情况进行预测,并通过红、黄、绿三种颜色的信号灯对学习者的学习情况进行呈现^[14]。

学习分析技术为 MOOCs 学习者的形成性评价提供了便利。《教育辞典》中对形成性评价的定义为:以诊断为主的评价,重在评价对象的工作状态和过程,其特点是通过揭示存在的问题,反馈有关信息,以促进教学工作的改进。美国心理学家布鲁姆首次将形成性评价应用到学生的学习中^[15],在教育领域,形成性评价主要用于教师通过分析学习者在整个学习过程中连续性的行为表现,进而对学习者的学习进行评价。由于学习分析技术的发展,在以 MOOCs 为代表的在线学习领域,越来越重视对学习者在在线学习过程的评价,并采用形成性评价的方式。在当前在线学习领域中,应用学习分析技术进行形成性评价主要依据在线学习者学习过程中产生的各种学习行为。学习者在线学习过程中的学习行为可归纳为三类:教学资源交互、学习参与度、形成性测试,教学资源交互包括视频完成情况,学习资源下载情况等行为。学习参与度包括学习者出勤情况,点击率,登录次数,在线时间,论坛讨论情况,参与互评等行为;形成性测试包括学习者在学习过程中参加的测试行为,主要为单元作业测试及期末测试等计入学生综合考核成绩部分学习者得分。由于 MOOCs 学习者参与人数众多,MOOCs 教学大多侧重于学习内容的呈



现,而对学生交互及学习支持服务相对欠缺,这导致学习者的学习参与度不高,难以实现真正意义上的教学资源交互^[16-21]。由于学习者在线学习行为的复杂性,在 MOOCs 教学中难以真正使用教学资源交互和学习参与度学习行为对学习者的学习进行评价,现阶段对学习者的形成性评价仍然主要依据形成性测试中的成绩数据。

(二)测试难度系数研究

通过在中国知网中检索“测试难度”关键词的结果可知,目前,国内外学者对试题测试难度系数的研究多集中于计算机应用领域,多是对传统试题难度和在线考试试题难度的研究,鲜有对 MOOCs 形成性测试难度的研究。测试的目的是为了反映学习者的学习情况,难度系数是反映测试试题对被测试学习者难易程度的一个重要直观参数,难度系数确定的准确程度直接影响到对测试质量的判断,进而可能影响到对学习者的学习效果做出正确评价,因而对难度系数的准确测量计算显得尤为重要。关于试题难度系数的计算主要有专家主观预估、经典测量理论(Classic Test Theory, CTT)和项目反应理论(Item Response Theory, IRT)三种方法。专家主观预估是专家根据个人经验预估试题难度的方法,这种方法具有一定的主观性;经典测量理论将难度系数的范围定义为 $[0,1]$,其取值取决于全体测试者的得分率,得分率越高,难度系数越低^[22],这种方法能避免专家预估的主观性弊端,但是缺乏一定的灵活性,难度系数的大小会随被试能力的高低而有所不同;项目反应理论用项目反应函数表示考生正确作答的概率与考生能力间的关系,每个项目反应函数对应一道题目,项目特征曲线的拐点位置反映了该题的难度^[23],项目反应理论强调以复杂数学模型为核心,具有研究困难、应用的可能性低的弊端^[24]。

基于基本难度系数计算方法,相关学者进行了大量难度系数方面的研究。华南师范大学王明兰教授针对专家评估试题难度存在的问题,通过一定量的被试试题得分率采用数理统计理论,通过 Hotelling 统计量及最小二乘估计来验证专家给定的难度系数^[25],这种方法虽然在一定程度上保证了论证的可靠性,但是由于试题难度测量选用被试的能力知识水平不同,尚缺乏一定的客观性。林雪明提出试题难度系数数学模型的建立和实现方法,通过对试题的得分率、区分度的计算,反复测试保证试题难度客观性,避免了专家确定试题难度的局限性^[26],该

模型难度系数虽然可自动进行修正,使确定的试题难度系数更加可靠稳定,但是此模型依旧存在局限性,对试题难度的稳定性不能确定。马江山针对经典测量理论 CTT(又称真分数理论)和项目反应理论 IRT 两种策略理论之间的相关性,以 IRT 为基础,充分利用 CTT 的优点,将两者有机地融合提出题库建设中试题难度的修订和校准,将区分度与难度进行相关度检验,并引入信度来保证试题难度的稳定性^[27]。2008年,艾小伟提出难度修正算法,将难度系数分为理论难度系数和样本难度系数,运用统计学习的原理构建了一个学习器模型,可根据考试样本数据进行学习修正^[28],在此算法模型中,经验风险最小准则以及统计学习理论,根据考试样本数据对试题的失分率、区分度、信度的计算进行学习修正,解决了试题的优化命题问题,使试题难度系数具有客观性和科学性。在2010年,王梅在艾小伟研究基础上引入波动率对试题难度并进行表征^[29]。

通过对难度系数相关研究的梳理,发现关于难度系数的研究基本上都是在最基本计算方法的基础上不断对已有模型进行优化,研究中大多用失分率、区分度、信度以及波动率来表征试题难度系数。在研究梳理的过程中,笔者没有发现关于难度系数在 MOOCs 形成性测试中进行应用的研究。基于 MOOCs 形成性测试特点,本研究从学习分析的视角,借鉴试题难度系数研究中难度系数的计算方法,依据 MOOCs 学习者的形成性测试成绩,构建 MOOCs 形成性测试难度系数模型。

三、MOOCs 形成性测试难度系数学习模型

由于在线学习过程中形成性测试与在线试题测试的相似性,本研究认为,MOOCs 单元测试作业可以与在线考试中的试题进行类比,学习者在 MOOCs 中做作业测试的过程可看成学习者做试题的过程。借鉴已有难度系数研究中利用统计构造难度系数学习模型的方式,为自动化计算 MOOCs 中形成性测试的难度系数,提出如下解决方案。

(一)初始难度系数

对于在线试题难度系数计算前初始难度系数的确定,为了避免专家评定的主观性和项目反应理论模型的复杂性弊端,决定采用经典测量理论,以 MOOCs 中所有学习者形成性测试成绩为研究样本来对难度系数的初始值进行计算。

(二) 难度系数影响因素

MOOCs 形成性测试的难度系数需要通过对学习者在测试中的成绩进行分析得到, 本研究选取试题难度系数影响因素研究中最常用的失分率、区分度和信度来表征难度系数, 三个影响因素分别定义如下。

1. 失分率(L)

失分率是测试学习者的失分情况表征, 学习者测试成绩失分率高, 表示该课程形成性测试的难度系数大; 反之, 难度系数越小。失分率依赖于测试样本, 而难度系数与测试样本关系不大。为使失分率更准确表征难度系数, 常常需要分不同批次的学习者测试成绩进行失分率分析, 这需要引入累计失分率(\bar{L}), 累计失分率是指形成性测试中从第一批到第 k 批次测试所有学习者的失分率。

2. 区分度(D)

区分度是衡量测试质量的重要指标之一, 是编制测验、筛选题目的依据。区分度取值越大, 表示形成性测试区分度越好, 说明能很好对学习者的学习效果进行区分。测量学家伊贝尔对试题的区分度评定如表 1 所示。

表 1 伊贝尔试题区分度范围划分

D 的范围	$D \leq 0.4$	$0.3 \leq D \leq 0.39$	$0.2 \leq D \leq 0.29$	$D \leq 0.19$
D 的评定	测试区分度良好	测试区分度较好	测试区分度一般 需要修改	测试区分度不好 应考虑淘汰

3. 信度(R)

信度是用来衡量测试可靠性与稳定性的指标, 信度高的测试说明测试者成绩分数不易受偶然因素造成的误差的影响, 可以比较真实地反映学习者的实际学习水平。一般认为, $B < 0.5$ 的测试题信度较差, $0.5 \leq B \leq 0.8$ 的测试题信度良好, $B > 0.8$ 的测试题信度非常好。MOOCs 形成性测试难度系数各影响因素与难度系数的关系, 可总结如下表 2 所示。

表 2 试题难度系数影响因素

影响因素	表征意义	影响关系
失分率(L)	衡量学习者测试失分情况	失分率越高, 测试难度系数越大; 反之, 难度系数越小
区分度(D)	衡量测试对学习者的区分程度	区分度越高, 测试区分程度越高; 反之, 区分程度越低
信度(B)	衡量测试对学习者的可靠性与稳定性	信度越高, 测试越可靠稳定; 反之, 越不稳定不可靠

(三) 样本难度系数序列收敛性论证

Vapnik 等提出的样本统计学习理论^[30](Sample

Statistical Learning Theory, SSLT), 是针对小样本情况研究统计学习规律的理论, 并通过构造机器学习的框架实现对推广能力的控制。SSLT 是传统统计学的重要发展和补充, 为研究有限样本情况下难度系数计算提供了理论框架。本研究基于 SSLT, 认为样本难度系数 g 与影响因素失分率 L 、区分度 D 及信度 R 之间存在一定的未知依赖关系。对于此依赖关系进行估计, 需要在预测函数序列中求一个最优函数来使期望风险最小。经验风险最小化(Empirical Risk Minimization, ERM)理论中, 经验风险是训练样本错误率, 学习的目标在于使期望风险最小化, 但是由于训练样本有限, 期望风险无法计算, 因此, 在传统的学习方法中采用了所谓经验风险最小化(ERM)准则, 即用样本来定义经验风险, 通过对参数求经验风险来逐渐逼近理想的期望风险的最小值。

本研究采用统计学习中的经验风险最小化(ERM)准则对期望风险进行推导, 利用概率密度估计中的最大似然法, 可以证明用样本难度系数 g_i 的统计量去估计理论难度系数 G , 能使期望风险损失最小。由最大似然法通过求偏导可得出, 理论难度系数 G 等于样本难度系数均值, 样本难度系数是理论难度系数的无偏估计。样本难度系数序列 $\{g_k\}$ 的收敛性是准确计算理论难度系数 G 的关键, 考虑用历史样本难度系数的线性组合来对未来数据如 g_{k+n} 进行预测和估计。由 Hilbert 空间性质可知, 样本难度系数的基本序列必收敛于一个唯一确定的常数, 该常数即为课程形成性测试题的理论难度系数 G 。

(四) 难度系数机器学习模型

基于样本难度系数的收敛性以及难度系数序列满足 ERM 准则, 则可以通过有限小样本数据, 近似求得理论难度系数 G 。传统试题难度系数机器学习模型定义如下:

$$g_k = \frac{\sum_{i=0}^{k-1} g_i + \varphi(L_k)}{k+1}, k=1, 2, \dots \quad (1)$$

上式中 g_k 为第 k 批次学习者样本难度系数, 本研究认为难度系数最直接的表征因素是失分率, 在此基础上与区分度和信度存在一定关系, 在用三个影响因素表征难度系数时对区分度和信度分别用它们区分好坏等级的临界值 0.4 和 0.8 作为标准, 因此样本难度系数 $\varphi(L_k)$ 计算公式定义如下:

$$\varphi(L_k) = L_k * \frac{D_k}{0.4} * \frac{B_k}{0.8} \quad (2)$$

理论难度系数 G 取值范围为: $0 \leq G \leq 1$, 本研究

http://dej.zjtvu.edu.cn

认为对 MOOCs 形成性测试来说,难度系数合理的范围是[0.2,0.8],高于这个区间或者低于这个区间的难度系数,反映该课程形成性测试不理想,不能对学习者进行有效形成性评价。基于难度系数 0 到 1 之间的范围限制,如果 $\varphi(L_k) \geq 1$ 或者 $\varphi(L_k)=0$,令 $\varphi(L_k)=\bar{L}_k$ (累计失分率)。公式(1)(2)中 $g_k, L_k, \bar{L}_k, D_k, B_k$ 分别表示第 k 批次测试中样本难度系数、失分率、累计失分率、区分度和信度。在本研究中第 k 批次课程形成性测试中样本难度系数均值定义为:

$$G_k=(g_0+g_1+\dots+g_k)/(k+1),k=1,2,\dots \quad (3)$$

(五)难度系数自动计算流程

在 MOOCs 中,可以采用将学习者随机分组的形式,对不同批次的学习者在章节或周作业测试中的成绩情况进行分析,应用难度系数机器学习模型自动收敛,得到 MOOCs 形成性测试理论难度系数。具体计算流程归纳如下:

(1)获取研究样本:计算选取的样本需为 MOOCs 中所有参加形成性测试的在线学习者在形成性测试中的测试成绩。

(2)计算初始难度系数:统计学习的方法要求给定试题难度系数的初始值,应用经典测试理论的方法基于全部样本确定形成性测试初始难度系数,这样可以避免专家给定的主观性弊端。

(3)计算样本难度系数影响因素:将样本随机分成不同批次,依据样本的测试成绩计算不同批次的学习者成绩的失分率、区分度、信度。

(4)计算收敛难度系数:利用难度系数机器学习模型,通过各批次样本难度系数影响因素和初始难度系数,计算测试的样本难度系数,进而收敛出测试理论难度系数。

四、MOOCs 形成性测试难度判断案例分析

(一)样本选取

本研究选取 365 大学平台上一门典型 MOOC 《思想道德修养与法律基础》^[31]来对形成性测试难度系数进行应用研究。该课程于 2015 年 10 月 1 日正式开课,“学堂在线”上的学习者可以在开课期间自由加入该课程学习,也可以由任课教师或管理员来添加学习者。该课程教学内容包括绪论在内共九章内容,所有学习资源均为内容讲解视频,按照学习进度,教师会提前上传视频学习资源及测试作业题,课程包括绪论在内共 6 章 30 节 157 个视频,每节课后

对应一个作业,共 30 个作业,每个作业总分 100 分。该课程考核标准为:完成课程学习,课程习题之和大于 60 分,即为合格,学习者最后学习成绩是依据 30 个作业按照相同权重取平均进行计算。该课程后台数据记录选课人数共 5410 人,剔除退课的 10 人及没有做过作业的 384 人后,共剩余 5016 名学习者样本。本研究选取 5016 名有测试作业行为的学习者的 30 个平时作业的成绩作为研究样本,将样本随机分为 10 组,平均每组 500 左右作为形成性测试的 10 个批次样本。

(二)计算初始难度系数

研究样本中有 4754 学习者最终考核合格,262 名学习者最终考核不合格,考核通过率约为 94.8%。课程通过率较高可能与该 MOOC 学习者中有近一半的学习者是某大学学生,该门课对他们来说是必修课程有关。采用经典测试理论的方法,以所有参与测试作业学习者的作业成绩为样本,计算得出该课程形成性测试初始难度系数 $g_0=0.11$ 。

(三)计算样本难度系数影响因素

分别计算案例课程 10 个批次测试难度系数影响因素失分率及累计失分率、区分度、信度,结果如表 3 所示。

表 3 案例难度系数影响因素计算结果

测试批次 k	测试人数 n	失分率 L	累计失分率 \bar{L}	区分度 D	信度 B
1	517	0.115	0.115	0.091	0.971
2	502	0.115	0.115	0.118	0.972
3	495	0.100	0.110	0.089	0.964
4	495	0.106	0.109	0.061	0.962
5	501	0.113	0.110	0.101	0.972
6	512	0.107	0.109	0.112	0.962
7	502	0.105	0.109	0.117	0.965
8	509	0.104	0.108	0.090	0.965
9	502	0.116	0.109	0.067	0.970
10	481	0.118	0.110	0.092	0.969

(四)计算收敛难度系数

依据难度系数机器学习模型用 MATLAB 软件编程得到 10 批次测试样本难度系数 g_k 和样本难度系数均值 G_k 如表 4 所示,两者收敛趋势图如图 1 所示。

在图 1 中,纵坐标表示难度系数,横坐标代表测试批次,图中上部蓝线表示样本难度系数均值,下部红线表示样本难度系数。可以看出,随着测试批次的增加,虽然样本不是很多,但是样本难度系数和样本难度系数均值序列收敛趋势较为明显,并且样本难

表 4 样本难度系数与样本难度系数均值结果

测试批次 k	样本难度系数 g_k	样本难度系数均值 G_k
1	0.0709	0.0904
2	0.074	0.085
3	0.0704	0.0813
4	0.069	0.0789
5	0.0715	0.0776
6	0.0717	0.0768
7	0.0718	0.0762
8	0.0708	0.0756
9	0.0704	0.075
10	0.0712	0.0747

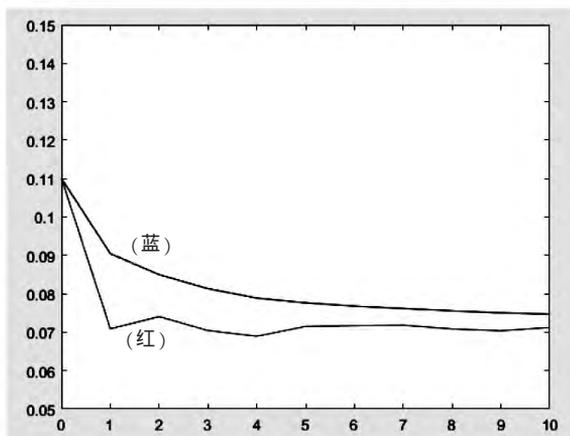


图 1 样本难度系数收敛情况

度系数均值收敛速度较快且比较稳定, 收敛的常数 0.074, 可以近似看作该课程形成性测试的理论难度系数。

五、案例结论探讨

关于试题难度系数的计算研究大多是在已有成熟算法基础上进行的改进, 本研究计算 MOOCs 形成性测试试题难度系数模型, 应用的是在统计分析领域计算试题难度系数较为成熟的小样本收敛算法, 计算数据采用能直观体现学习者测试情况的测试得分, 方法应用依据 MOOCs 形成性测试和传统测试的相似性, 本身具有一定的科学性。在案例课程应用中, 建立的模型可以依据小样本(共 10 批次)学习者测试成绩收敛出较为合理的理论难度系数, 该模型的可行性得到了实际的验证。

案例课程形成性测试根据总样本计算初始难度系数为 0.11, 通过模型收敛得出形成性测试理论难度系数大致为 0.074, 这远远超出难度系数合理范围, 因此, 可以据此判断该门课程形成性测试过于简单, 学习者很容易得到高分, 不适合用来对学习者进行形成性评价。

(一) 不同批次样本难度系数影响因素之间存在较小的差异

将案例课程学习者样本随机分为 10 组分别计算各组难度系数影响因素失分率、区分度和信度, 发现虽然是随机分组该案例课程, 在失分率和信度上分别有两组和六组在该影响因素上两两相同, 比如, 第 1 批次和第 2 批次失分率都为 0.115, 第 2 批次和第 5 批次信度都为 0.972。所有批次在三个影响因素上均存在较小的差异, 说明该案例课程学习者形成性测试整体比较稳定统一。

(二) 案例课程形成性测试比较稳定可靠, 区分程度较差

该案例课程形成性评价难度系数影响因素中的区分度在 10 批次测试中的平均值为 0.0938, 远小于 0.19, 说明该课程形成性测试区分程度很差, 应该考虑淘汰或改变该课程的形成性测试, 而在 10 批次测试中难度系数影响因素的信度平均值为 0.9672, 远大于 0.8, 说明该课程形成性测试比较稳定可靠。

(三) 案例课程形成性测试难度计算偏低

案例学习者整体合格率为 94.8%, 初始难度系数为 0.11, 失分率平均为 0.1099, 最后通过学习模型收敛出来理论难度系数约为 0.074。经过统计, 发现在 5016 名学习者中最终成绩 ≥ 80 分的共有 4178 人, 占 83.3%, 而 ≤ 20 分的只有 154 人。可以看出, 该课程中大部分学习者取得极高的成绩。本研究探求该案例课程形成性测试难度计算结果很低情况, 分析其原因存在以下几点。

第一, 该 MOOC 属于有学分的 MOOC, 并且学习者中有约 30% 的学习者是大学新生, 该 MOOC 对学习来说属于必修课程, 所以大部分学习者有较为明确的目标去取得好的测试成绩。在该 MOOC 中, 最终学习者成绩完全依据学习者平时作业测试成绩, 这会让大部分只是为了拿到学分的学习者为了取得好的成绩而仅仅关注与平时作业测试相关的学习内容, 这也在一定程度上反应了当前 MOOCs 学习中形成性评价依赖学习者成绩存在的弊端。

第二, 该 MOOC 形成性测试是以作业的形式布置给学习者, 测试的目的主要是为了保证学习者参与了学习, 形成性测试在这门课中变成督促学习者学习的手段而非真正意义上的测试。该 MOOC 测试题型大部分为选择题型, 学习者一般有多于一次提交的机会, 只要其中有一次选择正确后台即判断作答正确, 在这样在师生分离的情况下进行测试, 学生



较容易通过其它渠道(互联网搜索等)得出答案,也不排除有学习者通过注册多个账号进行作弊^[32]。

(四)MOOCs 中形成性评价现状反思及改进建议

案例课程难度系数计算结果极低的原因,反映了当前在大部分 MOOCs 中仅仅通过测试成绩对学习进行评价存在的问题。MOOCs 学习者大规模参与的特质,带来学习者学习需求的显著差异,MOOCs 平台上自动记录的学习者成绩,由于在线学习测试的复杂性不能完全用来反映学习者的学习效果。MOOCs 中的大规模学习者使得教师无法像传统教学那样,通过对学习者进行有针对性测试而获知学习者学习掌握情况。学习分析技术的发展提升了对 MOOCs 学习者学习效果评价及提供个性化学习支持的效果。但在现阶段,由于技术发展的相对不成熟性,学习分析能较好地处理对良构数据(以学习者成绩为主)进行处理分析,而对于非良构数据(如,学习者帖子内容等)的处理分析效果还不是很好。在本研究中,针对 MOOCs 形成性测试计算其难度系数的模型主要计算依据也是 MOOCs 平台记录的学习者测试作业成绩。基于现阶段应用学习分析技术进行形成性评价存在的上述困难,结合案例难度系数结果,认为可以通过以下两种途径改进 MOOCs 中形成性评价效果。

1. 对不同学习者进行有针对性分类测试评价

在 MOOCs 中动辄过万学习者的参与,使得教师难以对学习者的学习效果做出评价,现阶段 MOOCs 通过学习分析技术对学习者的良构数据进行分析进而做出判断是对学习者学习效果做出评价的一种有效方法,虽然用成绩来评价学习者的学习屡遭诟病,但是,在 MOOCs 中最直接有效的良构数据就是学习者的测试作业成绩。MOOCs 学习者学习需求差异显著,这导致学习者测试作业成绩的不可靠性。根据学习者的不同学习需求对学习进行分类,进而有针对性的对其进行测试,相似学习者测试的成绩会相对更加可靠,这样再依据成绩对学习进行分类进行评价,可能会提升 MOOCs 形成性评价的效果。

2. 探索多元化形成性测试评价手段

在线学习环境的复杂性导致 MOOCs 学习者学习成绩不可靠。本研究案例测试题型多为选择、判断等客观性题目,学习者可以很容易获得正确答案,这也导致根据学习者测试成绩计算出来的难度系数极低。在 MOOCs 形成性测试中,可以采取增加主观题比例等方法来有效降低学习者获取答案的容易程度。显然,学习者的测试成绩不能完全表征在线学习

者的学习效果,这跟传统学习一致。一个人的学习效果体现在学习过程中的方方面面,并且在很大程度上与学习者学习需求目标相关。对 MOOCs 学习者进行形成性评价,需要利用学习分析技术探索多元化评价手段,比如,通过学习者课程资源观看程度,课程学习内容交互情况等,都可以作为对学习形成性评价的参考指标。

六、总结

结合传统测试题与 MOOCs 形成性测试的一致相似性,引用统计学中难度系数小样本收敛方法,构建 MOOCs 形成性测试难度系数学习模型,该模型在实际案例 MOOC 中可以依据学习者的测试成绩在小样本情况下自动收敛出 MOOCs 形成性测试难度系数。本研究属于学习分析视角下基于学习者结构化数据(成绩)进行学习评价的一种探索尝试,在 MOOCs 环境下,学习者形成性测试情况比较复杂,成绩不能完全反应学习者的真实学习水平。难度系数的计算完全基于学习者形成性测试成绩,因此,本研究适用的课程领域须具有如下两个特点:第一,该课程学习者的形成性测试成绩需要相对真实客观;第二,该课程对学习者的形成性评价完全基于形成性测试成绩。

虽然,依靠 MOOCs 形成性测试成绩进行难度系数的收敛计算存在一定的不足,但是并非为了固化形成特定的数据标准,更多的是为在 MOOCs 形成性测试现状下为教师或管理者提供直接方便的难易程度反馈。后续会在以下两方面对 MOOCs 形成性测试难度系数继续深入研究:(1)加入成绩以外的其它影响因素(比如,测试前学习者观看学习相应课程资源情况等)来更加客观全面的反映 MOOCs 形成性测试的难易程度;(2)通过分析学习者做测试题的在线行为(比如,测试题提交前保存次数等)配合学习者该测试题得分,来计算分析在 MOOCs 中每次形成性作业测试的难度系数。

[参考文献]

- [1]徐思源.互联网+教育=?[J].基础教育课程,2015(13):74.
- [2]张喜艳,王美月,高嵩.教育网络信息生态系统生态化程度测评与优化[J].中国电化教育,2015(10):68-74.
- [3]郑勤华,李秋菊,陈丽.中国 MOOCs 教学模式调查研究[J].开放教育研究,2015(6):71-79.
- [4][7]金慧,刘迪,高玲慧,宋蕾.新媒体联盟《地平线报告》(2016 高等教育版)解读与启示[J].远程教育杂志,2016(2):3-10.

- [5]陈向东,曹安琪.为什么没有坚持——一个 MOOC 学习个案的分析[J].现代远程教育,2014(2):9-14.
- [6]王萍.大规模在线开放课程的新发展与应用:从 cMOOC 到 xMOOC[J].现代远程教育研究,2013(3):13-19.
- [8]顾小清,张进良,蔡慧英.学习分析:正在浮现中的数据技术[J].远程教育杂志,2012,30(1):18-25.
- [9]徐鹏,王以宁,刘艳华,等.大数据视角分析学习变革——美国《通过教育数据挖掘和学习分析促进教与学》报告解读及启示[J].远程教育杂志,2013(6):11-17.
- [10]张羽,李越.基于 MOOCs 大数据的学习分析和教育测量介绍[J].清华大学教育研究,2013,34(4):22-26.
- [11]郁晓华,顾小清.开放教育下的学习分析——2015AECT 夏季研讨会评述与延伸[J].远程教育杂志,2015,33(5):14-23.
- [12]武法提,牟智佳.基于学习者个人行为分析的学习结果预测框架设计研究[J].中国电化教育,2016(1):41-48.
- [13]余胜泉.学习资源建设发展大趋势(上)[J].中国教育信息化·高教职教,2014(1):3-7.
- [14]吴永和,曹盼,邢万里,等.学习分析技术的发展和挑战——第四届学习分析与知识国际会议评析[J].开放教育研究,2014(6):72-80.
- [15]Educationalinnovation C F. Formative assessment: improving learning in secondary classrooms[J]. Future Survey, 2005.
- [16]郑志高,张立国,张春荣.xMOOC 的学习评价方法调查研究[J].中国电化教育,2014(11):44-49.
- [17]曹中一,朱颖,杨静.网络教育质量与质量保障探索[J].创新与创业教育,2015(4):65-69.
- [18]朴杰,杨琳丽,曹德品.关于形成性评价的思考[J].中国高等医学教育,2010(5):4-5.
- [19]张建华.在线形成性评价的有效性研究[D].北京:首都师范大学,2007.
- [20]叶洪波.现代远程教育学习形成性评价研究与评价系统实现[D].长沙:中南大学,2008.
- [21]谢浩,陈丽.现代远程教育中形成性评价的现状及对策研究[J].现代远程教育研究,2010(4):49-53.
- [22]BORGATTO A F, AZEVEDO C, PINHEIRO A A. Comparison of ability estimation methods using IRT for tests with different degrees of difficulty [J]. Communications in Statistics-Simulation and Computation, 2015, 44(2): 474-488.
- [23]杨涛,辛涛,杨婷婷.试题难度的主观预估方法[J].中国考试,2014(2):3-9.
- [24]俞晓琳.项目反应理论与经典测验理论之比较[J].南京师大学报:社会科学版,1998(4):79-82.
- [25]王明兰,叶恒青.评估试题难度的检验[J].华南师范大学学报:自然科学版,2001(2):55-58.
- [26]林雪明.试题难度系数确定数学模型的建立与实现[J].杭州应用工程技术学院学报,2001,13(1):47-51.
- [27]马江山.题库建设中试题难度的修订和校准[J].数学的实践与认识,2006,36(9):276-280.
- [28]艾小伟,王有远.基于统计学习的试题难度系数修正算法[J].计算机工程与应用,2008,44(36):227-229.
- [29]王梅.基于样本分析的试题难度系数修正算法[J].西安邮电学院学报,2010,15(6):163-166.
- [30][美]VLADIMIR N V. 统计学习理论[M].北京:电子工业出版社,2009.
- [31]《思想道德修养与法律基础》课程详情页 [EB/OL].[2016-01-20]. http://xuetang.365dx.com/courses/course -v1:365DX +10610183X +2015_T2/about.

[作者简介]

宗阳,北京师范大学远程教育研究中心硕士研究生,主要研究方向:远程教育、在线学习分析;郑勤华,北京师范大学远程教育研究中心副教授,管理学博士,硕士生导师,主要研究方向:教育经济学、学习分析;张玄,北京师范大学远程教育研究中心硕士研究生,主要研究方向:在线学习分析;陈丽,北京师范大学远程教育研究中心教授,理学博士,博士生导师,主要研究方向:远程教育、终身教育。

The Research about Difficulty Coefficient of MOOCs' Formative Tests in the Perspective of Learning Analytics

Zong Yang, Zheng Qinhu, Zhang Xuan & Chen Li

(Research Center of Distance Education, Beijing Normal University, Beijing 100875)

[Abstract] Along with the development of learning analytics technology, we pay more and more attention on the MOOCs learners' learning process and usually use formative assessment method to evaluate learners' learning outcome. The tests of appropriate difficulty in online assessments can help learners to surpass their zone of proximal development and then arouse their learning enthusiasm. Learners' final grades in MOOCs are mostly decided by their scores of formative and final tests which combine together according to a certain weight proportion. Formative assessment of learners now mainly depends on their formative test grades. So how difficult are the formative tests in MOOCs? How to analyze the difficulty coefficient of MOOCs' formative tests based on learners' formative test score in the perspective of learning analytics? Based on the similarity of formative tests and traditional tests, the statistical method of small sample convergence which is usually used to calculate traditional tests' difficulty coefficient can be used to build the automatic learning model of MOOCs' formative tests' difficulty coefficient calculation. This model gets scientific and feasibility validation in a Chinese MOOC. The research method of MOOCs' formative tests' difficulty coefficient calculation can provide certain reference when setting formative tests.

[Keywords] Learning analytics; MOOCs; Formative tests; Difficulty coefficient

收稿日期:2016年3月22日

责任编辑:陈媛