

基于项目反应理论的学生能力测评研究

Research on Student Ability Evaluation Based on Item Response Theory

李志军 徐锦堂 (北方工业大学电气与控制工程学院, 北京 100144)

摘要:能力测评是测量学生各项能力的重要手段,通过能力测评能够全方位地了解学生的能力,从而进行有效指导。以某高校“自动控制原理”课程为例,在2020年春季疫情期间全程进行线上教学,并收集全年级学生测试结果,通过项目反应理论双参数逻辑斯蒂模型,计算出学生的各项相关能力,以便针对学生的能力短板进行精准化教学。

关键词:项目反应理论;能力测评;教育测量学

Abstract: Ability assessment is an important means to measure the various abilities of students. Through ability assessment, students' abilities can be fully understood and effective guidance can be given. In this study, taking the "Principles of Automatic Control" in a university as an example, online teaching was carried out during the spring 2020 epidemic, and the test results of the students of the whole year are collected. Through the project response theory two-parameter logistic model, the students' various Relevant abilities in order to accurately teach students based on their shortcomings.

Keywords: item response theory, ability assessment, educational measurement

随着大数据、人工智能等信息技术的不断发展,新型教学模式开始出现与兴起,传统的教育模式正处于改革中,从教育理念到管理机制、培养目标到评价体系,到教学内容到课堂形式,都面临着前所未有的挑战。国内的教育仍然延续传统的教学方式,并未做到真正的“因材施教”。以能力输出为目标的培养方式逐渐崭露头角,出现在人们的视野当中。

能力培养的核心思想在于提升学生的心理认知水平,从而影响学生在不同领域和环境下的学习活动效率^[1]。目前学术界对于能力的定义以及能力的测量与评价各执己见,未成体系。牛顿平台(Knewton Platform)开创了教育大数据个性化服务设计和应用的先河,通过采集学生的在线学习数据,可精准分析和预测学生的优势、不足、学习兴趣、认知投入水平,但其对于能力测评并未给出具体标准化方案^[2]。国内教育平台“松鼠AI”系统,搭载可测量、可定义的能力体系,能够快速检测被试者的学习能力,是国内比较领先与完善的能力评价体系。但其能力体系主要面向K12领域,整体性和迁移性不足^[3]。

本文以高等教育工科学科为背景,在疫情期间全程进行线上教学和测试,收集了全年级200多名学生多次测试结果,通过教育测量学中的项目反应理论(IRT),对不同学生各个维度的能力进行测量与评价,最终将学生能力进行量化,有利于直观地分析不同学生的能力差异,以便针对学生的能力短板进行精准化教学。

1 项目反应理论

项目反应理论^[4](Item Response Theory,以下简称IRT)是教育领域比较常用的心理测量理论,被测者某项特征的值主要是通过分析测试过程中对问题的回答情况来得到的,其意义在于可以指导项目筛选和测验。项目反应理论假设被试有一种“潜在特质”,潜在特质是在观察分析测验反应基础上提出的一种统计构想,在测验中,潜在特质一般是指潜在的能力,并经常用测验总分作为这种潜力的估算。项目反应理论认为被试在测验项目上的反应和成绩与他们的潜在特质有特殊的联系。通过项目反应理论建立的项目参数具有持久性的特点,意味着不同测量量表的分数可以统一。

IRT的模型有Logistic模型、Rash模型、Lord的正态卵形曲线模型等二十余种^[5]。Logistic模型是项目反应理论中常用的数学模型,其优势在于测试者对项目可能作出的定性反应的清

晰简洁的描述,是衡量二项计分题或客观问题更合适的模式。

假设在测试过程中只考察被测者的一种能力或潜质,那么被测者对测试的反应的概率可以描述如下:

$$p(\theta) = 1 - e^{-Da(\theta-b)} / [1 + e^{-Da(\theta-b)}] \quad (1)$$

该模型是Logistic双参数模型,其中, $D=1.702$, θ 表示被测者的能力估计值, a 表示区分度, b 表示难度, $p(\theta)$ 表示能力为 θ 的人答对此题目的概率。特征函数图像如图1所示。

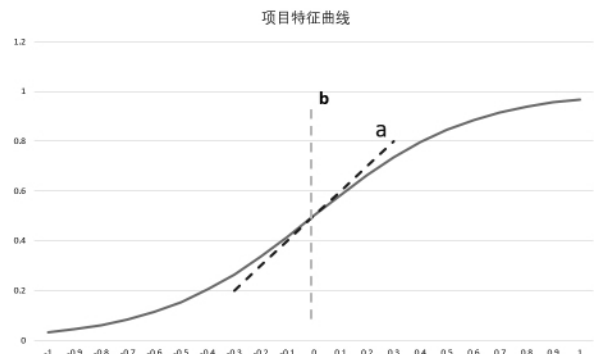


图1 项目特征曲线

2 参数估计

参数估计是应用IRT的前提,估计方法有多种,如条件极大似然估计、联合极大似然估计、边际极大似然估计等^[6]。本文采用条件极大似然估计法,其最适合用来估计Logistic模型IRT的估计值^[7]。

将试题相关参数代入项目反应理论中,得到其似然函数为:

$$L(x_1, x_2, x_3, \dots, x_n | \theta) = \prod_{i=1}^n P_i(\theta)^{x_i} Q_i(\theta)^{1-x_i} \quad (2)$$

其中, x_i 表示被测者对第*i*题的反应情况,如果答对则为1,答错则为0。

将(3)式两边取对数,得:

$$\ln L(x_1, x_2, x_3, \dots, x_n | \theta) = \sum_{i=1}^n [x_i \ln P_i(\theta) + (1-x_i) \ln Q_i(\theta)] \quad (3)$$

对上式求偏导,就能得出本次测试中被测学生该项学科能力的数值。

3 试题标化

根据项目反应理论的参数模型,需要对题库每一道试题进

行标注,试题标注由本校工科专业一线教师完成。部分试题标注如表 1 所示:

表 1 部分试题标注结果

试卷和序号	试题编号	考察能力	a	b
ZKCS1.1-1	ZKKU1-001	概念识别能力	0.37	0.26
ZKCS1.1-2	ZKKU1-002	关联辨析能力	0.41	0.25
ZKCS1.1-3	ZKKU1-003	关联辨析能力	0.52	0.56
ZKCS1.1-4	ZKKU1-004	概念识别能力	0.48	0.29
ZKCS1.1-5	ZKKU1-005	解释理解能力	0.69	0.49
ZKCS1.1-6	ZKKU1-006	直接应用能力	0.32	0.21
ZKCS1.1-7	ZKKU1-007	探索研究能力	0.54	0.24
ZKCS1.1-8	ZKKU1-008	直接应用能力	0.47	0.22
ZKCS1.1-9	ZKKU1-009	关联辨析能力	0.60	0.38
ZKCS1.1-10	ZKKU1-010	探索研究能力	0.49	0.27

本文根据课程每一个章节的内容设置 10 道试题,每一道试题考察学生的不同能力。a 表示试题区分度,区分度是指一个测验题目能够在多大程度上区分所要测量的潜在特质和能力,反映了测验题目对潜在特质和能力区分的有效性;b 表示试题难度,难度反映试题的难易程度,也可以理解为“容易度系数”。

根据布卢姆教育目标分类理论^[8],本文试题将着重考察学生五个关键能力^[9]:即“直接应用能力”“概念识别能力”“关联辨析能力”“解释理解能力”“探索研究能力”。其中“直接应用能力”和“概念识别能力”属于低阶水平;“关联辨析能力”和“解释理解能力”属于中阶水平之间;“探索研究能力”则属于高阶水平。

4 数据收集与处理

2020 年春季疫情期间,某高校工科专业“自动控制原理”课程全程进行线上教学和测试,收集了全年级 227 名学生共 14 次测试结果,并进行整理和处理,试题做对为 1,做错为 0。部分测试结果如表 2 所示:

表 2 部分学生测试结果

学生编号	1	2	3	4	5	6	7	8	9	10
1	1	1	0	1	0	0	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	0	1	1	1	1	0	0	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	0	1	1
6	1	1	1	0	1	1	1	0	1	1
7	1	0	1	1	0	1	1	1	1	1
8	1	1	0	1	1	1	1	1	1	1
9	0	0	0	1	0	1	0	0	1	0
10	1	1	1	0	1	1	0	1	1	1

5 学生能力测评结果

本文采集某高校工科专业“自动控制原理”课程多次测试结果,并进行整理和处理。通过教育领域常用的心理测量理论——项目反应理论对学生潜在特质和能力进行测量和评价,以便直观地了解每一个学生的各项能力。本次测试以第一章为例,计算得到部分学生相关能力值(如表 3)。

由测量结果可知,学生 2 和学生 7 各项能力都比较好,没有

表 3 部分学生能力测量值

学生	直接应用能力	概念识别能力	解释理解能力	关联辨析能力	探索研究能力
1	0.55	0.65	0.54	0.14	0.44
2	0.78	0.61	0.72	0.68	0.58
3	0.26	0.69	0.55	0.89	0.63
4	0.40	0.71	0.23	0.71	0.56
5	0.66	0.58	0.49	0.31	0.29
6	0.77	0.81	0.46	0.54	0.24
7	0.86	0.75	0.81	0.73	0.69
8	0.69	0.22	0.76	0.53	0.68
9	0.21	0.32	0.49	0.33	0.51
10	0.66	0.48	0.51	0.61	0.26

明显短板。学生 1 的关联辨析能力、学生 3 的直接应用能力、学生 4 的解释理解能力、学生 5 的关联辨析能力和探索研究能力、学生 6 的探索研究能力、学生 8 的概念识别能力、学生 9 的直接应用能力和概念识别能力,以及学生 10 的概念识别能力均存在不足,需进行针对性指导。后续可通过资源推荐系统为每个学生的能力短板进行针对性的资源推荐,从而提高学生能力。

6 结束语

本文通过项目反应理论双参数 Logistic 模型对某高校工科专业“自动控制原理”课程全年级学生进行多次测试,得到每个学生各个方面能力的测量结果,能够对学生的各项能力进行直观准确了解,以便对学生的能力短板进行针对性弥补,做到“因材施教”。能力测评的最终目的是希望能够根据不同学生的学业情况和能力实现精准教学,从而提升学生能力和成绩。

参考文献

[1]刘晨曦,郭二廓.新工科视野下经济类学生工程认知能力培养的教学探索[J].高教学刊,2019(8):78-83

[2]万海鹏,汪丹.基于大数据的牛顿平台自适应学习机制分析[J].现代教育技术,2016(5):5-11

[3]崔炜,薛镇.松鼠 AI 自适应学习系统[J].机器人产业,2019(4):84-94

[4]丁树良,罗芬,涂冬波.项目反应理论新进展专题研究[M].北京:北京师范大学出版社,2012

[5]陈冠宇,陈平.解释性项目反应理论模型:理论与应用[J].心理科学进展,2019(27):1-14

[6]朱樑,姚定康,梅长林,等.关于项目反应理论试题参数的估计[J].教育发展研究,2004,24(6):101-103

[7]Park Ryoungsun,Keenan A Pituch,Jiseon Kim,et al.Marginalized Maximum Likelihood Estimation for the 1PL-AG IRT Model[J].Applied Psychological Measurement 2015(4):448-464

[8]LW 安德森.学习、教学和评估的分类学——布卢姆教育目标分类学修订版(简缩本)[M].皮连生,主译.上海:华东师范大学出版社,2008:56-80

[9]申燕.SOLO 分类与关键能力相融合的学生思维评价——以 2019 年深圳市盐田区初中化学模拟考试为例 [J]. 教育测量与评价,2020 (5):42-48,55

[收稿日期:2020.12.17]

(上接第 121 页)
结构设计知识的调用与重用,对于提升设计效率具有重要意义。

本文与以往从无到有构建知识库不同,直接在原有 PDM 的基础上,加入了实例(课题)的属性参数库以及实例库(课题)的相应检索功能与推理功能,构建了一款可应用于企业内网的军用电子产品整机结构设计知识库。该知识库构建起来便捷可靠,成本较低,对原 PDM 的数据库可进行充分利用;且增加的数据量较小,基本不影响原有 PDM 系统的运行效率。

设计师可根据设计需求参数,在该知识库系统中进行相应检索与推理,得到符合该设计需求的实例课题的信息,这些信息

便于设计师更加快速地进行方案设计,从而提高整个团队的设计效率。

参考文献

[1]李德容,尤向华.面向军用电子设备的结构设计知识库[J].制造业信息化,2017,44(12):58-62

[2]梅启元.雷达结构设计智能化研究[J].现代雷达,2019,41(6):86-90

[3]张小勇.面向电子产品结构设计的知识库系统研究与开发[D].成都:电子科技大学,2012

[收稿日期:2020.12.11]