

新课程改革下高考物理选做题目等值性研究^{*}

——以重庆物理高考试题结构为例

冯庆^{1,2}, 廖伯琴¹, 洪宁², 李洪波³

(1. 西南大学 科学教育研究中心, 重庆 400715; 2. 重庆师范大学 物理与电子工程学院, 重庆 401331;
3. 重庆市教育考试院, 重庆 401147)

摘要:我国高中新课程改革物理学科分为必修4个模块(物理1, 物理2, 3-1和3-2)和选修4个模块(3-3, 3-4, 3-5和2-2)内容。在高考中,针对选修模块,国家课程标准卷和某些省市自主命题的试卷出现了选做题型,即不同的考试人群可以选择不同的试题作答以获取相应的分数,这必然引起试题得分的非等值性。研究以重庆市物理高考试题结构为例,通过百分位等值和线性等值法研究物理高考中选做题目的等值问题,得出选做题目分数不等值的定量结果,并提出修正选做题不等值的方法,为更好实现物理高考中选做题的公平性提供一种科学的判断依据。

关键词:教育测量与评价;新课程改革;物理高考;选做板块;试题等值性

中图分类号:G632.0

文献标志码:A

文章编号:1672-6693(2015)02-0168-05

高中新课程改革带来了物理新高考,高中新课程改革把培养高中生的创新精神、实践能力和健全人格作为基本目标。新课标教科书分为必修和选修,增加了学生的选择性,满足了不同学生不同层次的需要。对应新课标的选拔性考试是以能力立意为主的新高考,因此新课程标准下高考物理考试大纲对物理能力的考查提出了具体要求。新课程的多样性和选择性在课改实验区逐步体现,而对应的高考试卷则出现了针对选择模块的选做题。在2012年,全国有14个省市物理高考使用的是国家课程标准卷,国家课程标准试卷使用了选做题的考核方式。而在自主命题的省市中,重庆、山东、福建等省市也使用了选做题的形式。为了保证评价的公平性,命题时必须保证不同模块的选做题的难度和区分度基本一致,但是即使命题者主观认为不同模块的选做题难度和区分度基本一致,实际上的考试结果也会与预期值有一定的偏差,也就是说不同的试题对不同的选考群体一定会出现分数不等值的情况,这一难题必须运用“等值”技术解决考试中公平性问题。本文以重庆市即将实施的物理高考试题结构为例,通过百分位等值和线性等值方法提出了一种可供参考的计算和定量修正选做题非等值性的方法。

1 研究对象

重庆市教育考试院对重庆市的巴蜀中学、重庆第一中学、重庆第八中学、南开中学、松树桥中学、彭水中学、江津中学、渝北中学、重庆第十一中学、育才中学、重庆外国语学校、巴县中学和西南大学附属中学等13所中学进行了一次新课程物理高考的测试。在每所学校随机抽取两组学生,这两组学生的平均分成绩是相同的(即两组学生多次平时考试的综合平均分相同),然后让两组学生分别做两道不同的试题,试题分值均为12分,分别对应物理高考模块的3-3和3-4。由于两组学生平均分成绩是相同的,所以可以近似认为是两个水平相同的考试对象,可以排除考试对象的水平差异,直接研究试题本身的不等值问题。

两组学生测试试题的选做题目部分如下:选做题1对应3-3模块,主要考核热学内容,选做题2对应3-4模块,主要考核光学以及机械振动与机械波内容。重庆市的物理高考要求在3-3模块和3-4模块中选做其中一个模块的试题。

选做题1(试卷中的11题,对应3-3模块):

* 收稿日期:2013-11-10 修回日期:2014-11-12 网络出版时间:2015-01-22 11:56

资助项目:重庆市人文社会科学重点研究基地项目(No. 12SKB016);重庆市教育科学“十二五”规划课题(No. 2011-KS-023);国家博士后基金委项目(No. 2011M501376);重庆市教委科学技术研究项目(No. KJ120614)

作者简介:冯庆,男,教授,博士后,研究方向为物理课程与教学论,E-mail: fengq_126@163.com

网络出版地址: <http://www.cnki.net/kcms/detail/50.1165.n.20150122.1156.020.html>

1)(5分)分子间的作用力 F 与距离 r 的关系如题 11 图(图 1)所示。若有一分子 P 从无穷远处以一定的初速度沿 PQ 连线向固定不动的分子 Q 靠近, P 、 Q 间作用力

- A. 大小逐渐减小, P 在 r_0 处动能为零。
- B. 大小逐渐增大, P 在 r_0 处动能最大。
- C. 开始表现为斥力, P 过 r_0 后表现为引力。
- D. 开始表现为引力, P 过 r_0 后表现为斥力。

2)(7分)某实验室的空间体积为 V_0 ,室内压强为 p_0 ,气体温度为 $T_0 = 280$ K,通过空调升温后,室内温度变为 $T_1 = 300$ K,室内压强仍为 p_0 ,求升温前后该实验室内气体的质量之比。

选做题 2(试卷中的 12 题,对应 3-4 模块):

1)(5分)某同学做杨氏双缝干涉实验,用黄光照射双缝,屏上出现明暗相间的条纹。保持挡板与屏之间距离不变,增大两狭缝间距离,然后换用另一色光照射双缝,发现条纹间距恰好与黄光照射时的相同,可以判断这种色光可能是

- A. 红光 B. 绿光 C. 蓝-靛光 D. 紫光

2)(7分)一波长为 λ 的简谐横波在介质中沿 x 轴传播, P 、 Q 是 x 轴上的相距为 l 的两点, P 处的质点在波峰时, Q 处的质点刚好处于波谷,若 $\lambda < l$,求该波波长的最大可能是多少。

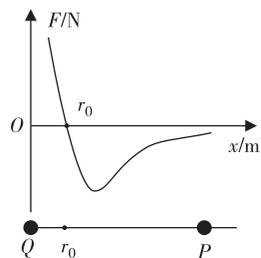


图 1

2 研究方法

2.1 难度分析方法

难度是被试完成题目时所遇到的困难程度,以全体被试得分率为难度系数,难度系数的计算公式^[4-5]:

$$P = \bar{X} / X_{\max}, \quad (1)$$

其中 P 代表题目难度, \bar{X} 为被试在某道题目上的平均得分, X_{\max} 为该题目的满分。用上述方法计算出来的难度系数属于顺序变量,即所测难度在数量大小基础上进行有序的分类,不具有相等的单位间隔,指出的仅仅是题目的相对难度。因为 P 的比较并不能客观指出题目难度之间差异大小,所以常假设每个试题所要测量的潜在特质或能力是呈正态分布的,并根据正态分布曲线,将试题的难度值 P 作为正态曲线下的面积,转换成具有相等单位间隔的等距量表,即用 Z 分数进行难度比较。本文采用 P 值和 Z 值作难度比较的参数。

2.2 区分度分析方法

区分度是水平能力的量度,常记为 D ,分析被试在测题上的得分与在整个测验总分之间的一致性程度。区分度的计算——积差相关系数,对于非二分法计分的项目,因得分具有连续性,在被试团体较大时可以认为题目分数服从正态分布。可将题目得分与测验总分求积差相关系数以得到题目的区分度,其计算公式^[1-2]为

$$r_{rx} = \frac{S_x^2 + S_y^2 - S_D^2}{2S_x S_y}, \quad (2)$$

式中 S_x 、 S_x^2 为全体被试在某题目上得分的标准差和方差, S_y 、 S_y^2 为全体被试总分的标准差和方差。 S_D^2 是由被试总分减去某题目得分获得一个新变量的方差。

2.3 等值测验

考生使用了不同的试卷或题目,但希望他们所得的分数仍然是可比的。这就需要通过一定的技术方法把它们关联起来,而采用的这一技术方法则称为测验等值^[3-5]。考生的考试分数不同有两方面的主要原因,其一是考生的学习能力不同,其二是试题不同。本文采用的测验等值^[6-7]的设计是单组设计,即对考生总体随即分组,分为组 1 和组 2,按统计理论可以认为两个组的考生的学习能力相等。排除考生学习能力不同这一因素,直接研究试题不同带来的等值性差异。组 1 的考生做测验 X,组 2 的考生做测验 Y。在两组考生分别完成了测验 X 和测验 Y 后,则可对两组考生的分数等值进行研究,包括线性等值和等百分位等值。

经典测验等值的方法分为等百分位等值和线性等值^[8-9],等百分位等值是认为两组相同学习水平的学生做总分相同的两道试题,相同百分等级的分数是等值的。例如:一道 12 分的试题,组 1 有 40% 的学生测试第 1 题得分低于 6 分,组 2 有 40% 的学生测试第 2 题得分低于 7 分,如果组 1 和组 2 学生的学习成绩是相同的,那么可认为第 1 题得 6 分与第 2 题得 7 分是等值的。一个分数的百分等级是指常模团体中低于该分数的人数百分比。百分等级的计算过程:先列出各原始分数的次数分布表,然后算出低于某一分数的人数(累计次数);将此人数除以样本总人数便可得到低于该分数的人数比例;此人数比例与 100 相乘所得结果即为百分等级。

线性等值^[1-2,10]的等值函数为

$$X^* = \mu_x + \frac{\sigma_x}{\sigma_y}(Y - \mu_y), \quad (3)$$

式中 μ_x 、 μ_y 、 σ_x 、 σ_y 分别是测验 X 和测验 Y 在总体上的平均值和方差。

3 研究数据分析与结果

3.1 百分等级等值处理

通过测量数据对 13 所中学选做题的难度和区分度参数分析如表 1 所示。

表 1 13 所学校的选做题 1 和选做题 2 难度值与区分度

学校	题目	难度值	区分度	被试人数/人	两题都未作答人数/人	总人数/人	学校	题目	难度值	区分度	被试人数/人	两题都未作答人数/人	总人数/人
重庆第十一中学	选做题 1	0.63	0.53	90	16	121	西南大学附属中学	选做题 1	0.69	0.33	54	42	112
	选做题 2	0.74	0.36	15				选做题 2	0.76	0.19	16		
巴蜀中学	选做题 1	0.7	0.43	88	24	121	重庆第一中学	选做题 1	0.78	0.61	49	36	119
	选做题 2	0.73	0.67	9				选做题 2	0.68	0.39	34		
巴县中学	选做题 1	0.6	0.28	45	19	121	渝北中学	选做题 1	0.51	0.24	22	14	68
	选做题 2	0.86	0.49	57				选做题 2	0.7	0.53	32		
江津中学	选做题 1	0.75	0.65	52	57	125	育才中学	选做题 1	0.78	0.34	65	8	89
	选做题 2	0.91	0.34	16				选做题 2	0.74	0.24	16		
南开中学	选做题 1	0.6	0.65	44	27	87	重庆第八中学	选做题 1	0.7	0.51	61	34	108
	选做题 2	0.64	0.26	16				选做题 2	0.71	0.2	13		
彭水中学	选做题 1	0.49	0.45	43	62	119	松树桥中学	选做题 1	0.58	0.43	54	36	111
	选做题 2	0.58	0.25	14				选做题 2	0.66	0.84	21		
重庆外国语中学	选做题 1	0.81	0.5	53	13	84	合计	选做题 1	0.68	0.53	720	388	1 385
	选做题 2	0.82	0.3	18				选做题 2	0.75	0.47	277		

经典测验等值的方法有等百位等值和线性等值两种。两道选做题目的百分位等级如表 2 所示。

从表 2 可以看出,大多数分数点上,题目 1 的百分位等级高于题目 2 的百分位等级,说明题目 1 是难于题目 2 的。例如在 10 分的分数点上,有 59.31% 的学生做题目 1 没有得到 10 分,而只有 47.65% 的学生做题目 2 没有得到 11 分。

将百分位等级换成分值 12 分,以选做题 1 为标准,将选做题 2 进行折合,如表 3 所示。

表 2 选做题 1 和选做题 2 的百分位等级表

原始分数 X	百分位等级(选做题 1)	百分位等级(选做题 2)	两道选做题的差值/%
12	65.00	51.26	13.74
11	64.03	49.82	14.21
10	59.31	47.65	11.65
9	54.58	44.77	9.82
8	53.61	43.32	10.29
7	42.50	27.08	15.42
6	39.86	26.35	13.51
5	4.72	3.97	0.75
4	2.92	3.97	1.05
3	2.08	3.61	1.53
2	0.83	1.08	0.25

表 3 百分位等级折合成分值 12 分的折合分数(以选做题 1 为标准)

选做题 1	选做题 2	相差的数值
12	9.46	2.54
11	8.56	2.44
10	8.03	1.97
9	7.38	1.62
8	6.46	1.54
7	4.46	2.54
6	3.97	2.03
5	4.21	0.79
4	5.44	-1.44
3	5.21	-2.21
2	2.60	-0.60

表3中得到折后不同分数的等值点,如做题1得12分与做题2得9.46分等值,做题得11分与做题2得8.56分等值,以此类推。

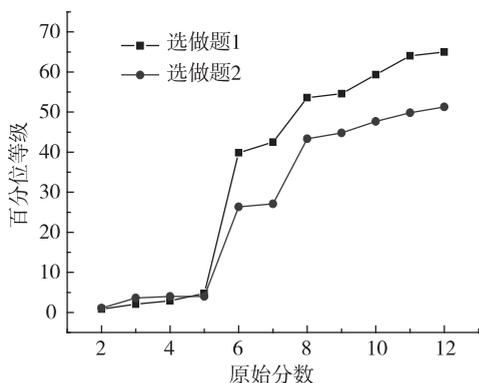


图2 两道选做题的百分位等级曲线

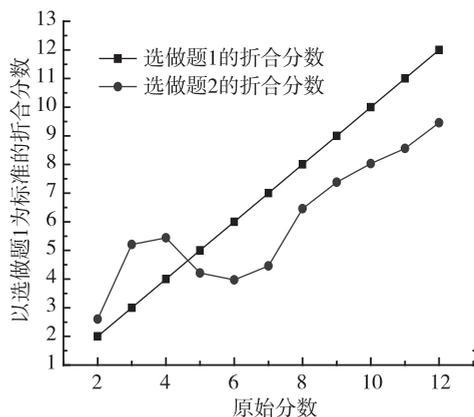


图3 以选做题1为标准的折合分数曲线

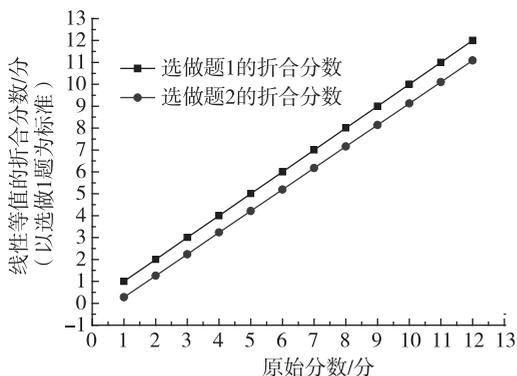


图4 两道选做题线性修正后的等值性对比

表4 该范例中线性等值函数的有关计算参数

选做题	期望	方差	标准差	难度值	区分度
选做题1	8.11	10.65	3.26	0.68	0.53
选做题2	8.97	10.84	3.29	0.75	0.47

表5 选做题2线性等值修正后的数值

选做题2的分值/分	选做题2修正后的分值/分	选做题1的分值/分	两道选做题的差值
12	12	12	0
11	10.10	11	0.90
10	9.12	10	0.88
9	8.14	9	0.86
8	7.16	8	0.84
7	6.17	7	0.83
6	5.19	6	0.81
5	4.21	5	0.79
4	3.23	4	0.77
3	2.24	3	0.76
2	1.26	2	0.74
1	0.28	1	0.72

从图2分析得:原始分数(即学生的得分)在2~6分(不包括6分)时两道选做题的百分位等级相差最大值为1.53%,但当原始分数(学生的得分)大于或等于6分时,两道选做题的百分位等级相差很大在9.82%~15.42%之间。将百分位等级换成分值12,如表3所示,这样可以和线性等值进行对比。从表3和图4可以看出,当得分在4分以下(含4分)时,选做题1的折合分数约低于选做题2,但是当得分在超过5分(含5分)时,选做题1的折合分数略高于选做题2,将上表的相差值相加计算两道选做题的平均差值是1分左右。

3.2 线性等值处理

两道选做题有关线性等值计算的计算结果如表4所示:

由线性等值可得公式

$$X^* = \mu_x + \frac{\sigma_x}{\sigma_y}(Y - \mu_y) \quad (4)$$

得到等值函数为

$$X^* = 8.11 + \frac{10.65}{10.84}(Y - 8.97)$$

通过表5和图4可以看出,两道选择题进行线性修正后的等值差别的平均分大约在0.8分左右,选做计算两道选做题的差值的平均值是0.8分。

4 结论

1) 通过数据分析可以得出在此研究对象下,选做题2要比选做题1简单,由等百分位等值图像可以看出在得分低于5分时,两道选做题是等值的。得分为5~12分的选做题2要修正,用百分位等值法给出的修正值是1分,即两道题目的不等值程度为1分;线性等值法给出的修正值是0.8分,即不等值程度为0.8分。

2) 用线性等值法得到的图像只能显示总体的性质,体现出分数等值差异的平均值(如图3),不能显示细节,比如不能比较两道选做题第一、第二问的难度。而百分位等值的方法较优,可以体现每个分值对应另外一道选做题中对应的分数。

3) 本项研究提出了两种修正物理高考选修模块不等值的方法,在实际考试中可以通过考试人群的抽样,通过对必修模块的作答分数,选

择出考试水平相当的两个群体,建立等值函数(包括百分等级函数和等值函数)的相关参数,在选修模块试题的等值性差异较大时进行等值性修正,为教育考试机构对物理选做试题的等值性测量提供了一种有效的定量衡量标准。

参考文献:

- [1] 黄光扬. 教育测量与评价[M]. 上海:华东师范大学出版社, 2002.
Huang G Y. Educational measurement and evaluation[M]. Shanghai:East China Normal University Press,2002.
- [2] 张敏强. 教育测量学[M]. 北京:人民教育出版社,1998.
Zhang M Q. Education measurement[M]. Beijing:People's Education Press,1998.
- [3] 马洪超. 基于项目反应理论的等值跨样本一致性的考察[J]. 中国考试,2011(11),16-20.
Ma H C. The investigation about the population in variance in IRT equating[J]. China Examination,2011(11),16-20.
- [4] 甘良梅,余嘉元. 标准参照测验分数体系的探讨研究 [J]. 心理学探新,2006,26(3):79-83.
Gan L M, Yu J Y. The study of criterion referenced tests score system[J]. Psychological Exploration,2006,26(3):79-83.
- [5] 任子朝,王蕾,朱乙艺,等. 标准参照考试理论在高考中的应用——以 H 省 2010—2012 年高考理科数学为例 [J]. 数学教育学报,2013,22(3),1-4.
Ren Z Z, Wang L, Zhu Y Y, et al. Application of criterion-referenced theory in the entrance examination to higher education evaluation——take 2010—2012 science mathematic results in H province for example[J]. Journal of Mathematics Education,2013,22(3),1-4.
- [6] 黎光明,张敏强. IRT 测验等值模型的选择——以广东佛山市中考数学实测数据为例[J]. 中国考试,2012(2):8-13.
Li G M, Zhang M Q, How to select equating models for college entrance examination based on item response theory [J]. China Examination,2012(2):8-13.
- [7] 张敏强,黎光明,焦璨. 普教“升中”考试中测验等值的应用研究——以广东省佛山市“升中”考试为例[J]. 心理与行为研究 2009,7(1):27-31.
Zhang M Q, Li G M, Jiao C. Application research of test equating in high school entering exam-exemplify high school entering exam of Foshan city in Guangdong province[J]. Studies of Psychology and Behavior,2009,7(1):27-31.
- [8] 丁树良,熊建华,毛萌萌. 项目反应理论框架下的新等值方法——对数对比等值法 [J]. 心理学报,2003,35(6):835-841.
Ding S L, Xiong J H, Mao M M. Log-contrast method for equating test based on IRT[J]. Acta Psychologica Sinica,2003,35(6):835-841.
- [9] 谢小庆. 考试分数等值的新框架[J]. 考试研究,2008,4(2):4-16.
Xie X Q. New framework in test equating[J]. Examinations Research,2008,4(2):4-16.
- [10] 谢小庆. 对 15 种测验等值方法的比较研究[J]. 心理学报,2000,32(2):217-223.
Xie X Q. Comparison of 15 equating methods[J]. Acta Psychologica Sinica,2000,32(2):217-223.

The Equivalence Study of Physical Selected Questions in College Entrance Examination under the New Curriculum Reform:

Take the Structure of Chongqing Physical College Entrance Examination as an Example

FENG Qing^{1,2}, LIAO Baiqin¹, HONG Ning², LI Hongbo³

(1. The Research Center of Science Education, Southwest University, Chongqing 400715;

2. College of Physics and Electronic Engineering, Chongqing Normal University, Chongqing 401331;

3. Chongqing Education Examination Authority, Chongqing 401147, China)

Abstract: Senior high school physics course is divided into 4 required modules (physics1, physics2, 3-1 and 3-2) and 4 elective modules(3-3, 3-4, 3-5 and 2-2) content under the new curriculum reform. In the entrance examination, for elective module, the national curriculum standard paper and some independent proposition provinces appeared optional questions. This means that different examinations crowd can choose different questions to answer in order to obtain the corresponding score, which would inevitably lead to non-equivalence of the exam score. In this paper, taking Chongqing physical college entrance examination structure as an example, through the percentile equivalent and linear equivalent method we study the equivalence of optional questions, and get the scores non-equivalent quantitative results of selected questions and propose amendments optional questions non-equivalent method. That would provide a scientific judgement to implement optional questions fairness to the physics college entrance examination.

Key words: educational measurement and evaluation; new curriculum reform; physical college entrance examination; optional modules; questions equivalence

(责任编辑 陈 琴)