

Rasch 模型在物理学业质量评价中的应用研究

肖月 桑芝芳

(苏州大学物理科学与技术学院 江苏 苏州 215000)

(收稿日期:2020-11-24)

摘要: Rasch 模型是估计题目难度和学生能力的一种数学模型,可以对测试试题难度的分布、学生的能力水平分布以及测试目标进行精确分析,为测试质量的分析提供一个新的视角.文章以一份高三物理试卷中客观题的质量分析为例,主要从整体质量检验、单维性检验、怀特图、试题拟合度、试题气泡图等方面介绍 Rasch 模型在物理学业质量评价中的应用.在具体的应用中,由于实际的测量情况不同,测量者应该结合测量目标,选择合理的 Rasch 模型对应的分析功能对测试进行分析.

关键词: Rasch 模型 物理试题 质量分析

1 Rasch 模型的原理

Rasch 模型是由丹麦数学与教育学家 G. Rasch 于 1960 年提出来的一个用来测量潜在特质的概率模型.它成功地解决了经典测量理论中被试样本依赖、测量工具依赖以及被试能力参数与项目难度参数之间相互干扰等问题.在教育与心理测量领域中,Rasch 模型常常依据学生在测试上的表现,间接测量出学生的潜在特质——学科能力.

1.1 Rasch 模型的数学表达式

在 Rasch 模型中,被试能力,试题难度,以及被

试给出正确答案的可能性之间的关系可以由方程式(1)来表达^[1].如下

$$p_{mi}(x_{mi} = \frac{1}{\theta_m}, \delta_i) = \frac{1 + \exp(\theta_m - \delta_i)}{\exp(\theta_m - \delta_i)} \quad (1)$$

其中 θ_m 表示被试能力, δ_i 表示试题难度, $p_{mi}(x_{mi} = \frac{1}{\theta_m}, \delta_i)$ 则表示答对概率.

1.2 Rasch 模型的参数估计

项目难度和被试的能力水平是 Rasch 模型中的两大参数.这两个参数的估计是 Rasch 分析中非常重要的过程.一般采用的估计方法为极大似然估计

5.3 强化解题的规范性训练 培养良好的解题习惯

规范解题包含规范的审题、规范的语言表述、规范的演算过程以及规范的解题结果^[4].高考物理计算题评分标准明确指出要有必要的文字说明、基本方程式和主要的解题步骤,只写最后结果的不得分.可见,学生规范性解题习惯的训练,能够有效地帮助学生减少失分,提高学生解题的正确率.

2020 年高考理综新课标卷Ⅲ第 25 题属于单体多过程问题,且不同过程涉及的位移、速度、加速度、摩擦力、时间等物理量均不同.学生在使用字母、符号时要注意:题中给定的字母意义要严格按照题目要求,常规通用的要与课本中字母表示的一致,不能随意设定,不同阶段的相同物理量之间要用不同的下标加以区分,并配以简洁的文字语言加以描述.

学生良好解题习惯的养成不是一蹴而就的,离不开教师和学生的共同努力.根据课程标准、考试大纲和学情现状,制定合理的习题教学方案,通过科学训练来促进良好解题习惯的养成.

参考文献

- 1 教育部考试中心.中国高考评价体系[M].北京:人民教育出版社,2019.8~9
- 2 封玮玮.追根求源 理解概念 拨开迷雾 揭示规律——基于学生认知规律的高中物理概念和规律教学改革实践研究[J].中国校外教育,2020(8):40~42
- 3 黄玉梅.拓展思维 一题多解[J].物理教师,2018,39(11):90~92
- 4 李晓辉.关于物理解题规范性的探究[J].教育教学论坛,2013(40):95~96

法,其原理为在已知被试作答矩阵的情况下,找到一个数值,使参数估计的可能概率达到最大值,通过求取似然概率的最大值来估计出考生的能力水平和项目的难度.

1.3 Rasch 模型与数据的拟合

Rasch 模型是一个理想化的数学模型.它要求所收集的实证数据必须满足事先规定的标准和结构,才能实现客观测量^[1].因此,Rasch 模型通过提供未加权残差均方(Outfit MNSQ)和加权残差均方(Infit MNSQ)两种形式的卡方拟合指标,来判断所收集的实证数据是否符合该模型.

本文采用了 Rasch 研究中经常使用的取值范围,即 Outfit MNSQ 和 Infit MNSQ 值在 0.5~1.5 范围内,说明拟合程度较好. Outfit ZSTD 和 Infit ZSTD 分别是 Outfit MNSQ 和 Infit MNSQ 的标准化的形式,Outfit ZSTD 和 Infit ZSTD 值在-2~2

范围内,说明拟合程度较好^[2]. 在分析过程中,如果有试题的拟合程度不理想,我们不应该直接删除该试题而是认真地审视该试题找出影响拟合度的因素^[3].

2 利用 Rasch 模型分析物理试题的质量

2.1 研究对象与研究方法

本研究对某高中 587 名高三学生在一份高三物理试卷中客观题的作答情况进行分析.采用 Excel 对收集到的原始数据进行处理和转化,然后将处理后的数据导入 WINSTEPS 软件,进行 Rasch 分析,其中主要包括整体质量检验、怀特图、单维性检验、试题拟合度以及试题气泡图.

2.2 试题的整体情况分析

试题的总体统计情况如表 1 所示.

表 1 试题的总体统计情况

	试题难度 (Measure)	Infit		Outfit		信度 (Reliability)	区分度 (Separation)
		MNSQ	ZSTD	MNSQ	ZSTD		
试题	0.00	1.00	-0.10	1.32	0.10	0.99	10.19

由表 1 可知,试题难度为 0.00;Infit MNSQ 和 Outfit MNSQ 分别为 1.00 和 1.32,Infit ZSTD 和 Outfit ZSTD 分别为-0.10 和 0.10,均在理想值范围内,说明本次数据与 Rasch 理想模型拟合程度较好;试题的信度(Reliability)为 0.99(>0.7),表示试题具有较高的可信度;试题的区分度(Separation)为 10.19(>2),显示试题能够对不同能力水平被试加以区分.

从内容来看,试题强化了对物理基本概念与规律等知识的考查,主要涉及匀变速运动、圆周运动、动能定理与机械能守恒、静电场等知识.突出考查学生对物理基本概念与规律的理解,有助于培养学生的物理观念与科学思维.而且每个试题都具有明确的考查目标和功能,能够对不同水平的学生进行区分.

因此,从定性和定量两个方面对客观题进行整体评估,结果显示整体情况较好,可进行后续分析.

2.3 学生能力与试题难度关系图(怀特图)

Rasch 模型将所收集的实证数据进行对数转换,使其变为等距数据(logit 数据).然后将学生能力和试题难度的等距数据标定在同一个量尺上,根

据它们在量尺上的相对位置,直接比较学生之间的能力高低,试题之间的难度大小,以及学生与试题之间的关系.

图 1 显示了学生能力和试题难度之间的对应关系.其中,左边部分表示学生能力的分布,每个#号代表 7 名学生,每个点表示 1~6 名学生,中间竖线是等距量尺,右边部分表示试题难度的分布.从上往下,学生能力逐渐降低,试题难度逐渐减小.

由图 1 可知,学生能力分布范围约为 5 个 logit,试题难度分布范围约为 6 个 logit.总体来看,学生能力平均值为 1.61 logit,考试题难度平均值为 0.00 logit,学生能力水平高于试题难度,说明试题难度偏低.试题难度没有覆盖到全体学生,特别是较高能力水平的学生.针对此类情况,需要适当增加一些难度较大的试题,并相应减少没有学生能力水平对应的难度较小的试题.

2.4 单维性检验分析

Rasch 模型的假设之一,测量的潜在特质是单维的,即学生在该测试中取得结果的好坏只与某一种潜在特质有关(如知识、能力、性格特征等),本文

通过 Rasch 模型提供的标准残差对比图,来鉴别该试题测量的潜在特质是否是单维的.

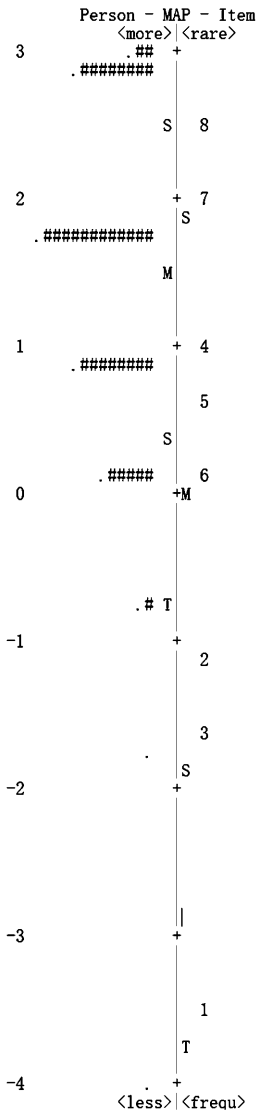


图1 怀特图

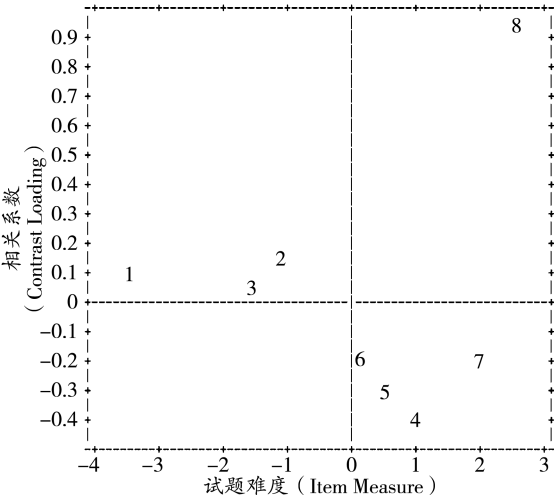


图2 标准残差对比图

图2展示了试题的标准残差对比图,横坐标是试题难度(Item Measure),左侧纵坐标是试题与其他可能的因素之间的相关系数(Contrast Loading),坐标系中的1~8为题号.可以看到,1~7题的相关系数都在理想范围 $[-0.4, 0.4]$ 内,具有单维性,以第3题为例进行说明,并对超出理想范围的第8题做进一步分析.

【第3题】如图3所示,车厢水平底板上放置质量为 M 的物块,物块上固定竖直轻杆.质量为 m 的球用细线系在杆上 O 点.当车厢在水平面上沿直线加速运动时,球和物块相对车厢静止,细线偏离竖直方向的角度为 θ ,此时车厢底板对物块的摩擦力为 F_f 、支持力为 F_N ,已知重力加速度为 g ,则()

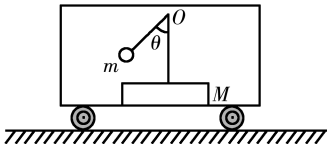


图3 第3题题图

- A. $F_f = Mg \sin \theta$
- B. $F_f = Mg \tan \theta$
- C. $F_N = (M + m)g$
- D. $F_N = Mg$

在标准残差对比图中,本试题与其他可能的因素之间的相关系数为0.04.它通过连接体模型考查牛顿运动定律的运用,学生先以小球为研究对象,由牛顿第二定律推导出车厢运动的加速度;然后以物块与小球整体为研究对象,通过力的分解求出水平方向与竖直方向上的力.在此过程中,并未涉及物理知识以外的其他潜在特质,具有单维性.

【第8题】如图4所示,在水平放置的光滑接地金属板中点正上方 h 高处,有一带正电的点电荷 Q ,一表面绝缘、带正电的小球(可视为质点,且不影响原电场)以速度 v_0 在金属板上自左端向右端运动,则()

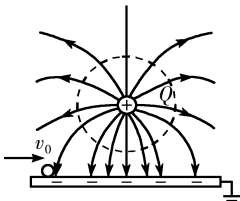


图4 第8题题图

- A. 小球先做减速后做加速运动
- B. 运动过程中小球的电势能先减小后增大
- C. 金属板上的感应电荷对点电荷 Q 的作用力

$$F=k\frac{Q^2}{4h^2}$$

D. 金属板上的感应电荷对点电荷 Q 的作用力

$$F=k\frac{Q^2}{h^2}$$

通过 Rasch 分析,第 8 题难度为 2.53 logit,只有 34% 的学生回答正确,是客观题中最难的题目,拟合度在理想范围内,能够区分不同能力水平的学生,区分度较好.该试题考查了静电平衡、力与运动以及功等知识点,综合性较强.要求学生掌握静电平衡导体特点,以及能根据题图电场线的分布,联想到等量异种电荷.如若由于思维定势,学生没有充分考虑到导体的放入,以及对之前所学知识不能情境迁移,则不能将题目信息与已有知识相匹配进而推断出正确答案.因此,除了学生的物理知识,本试题还考查了学生的推理分析、比较、情境迁移等综合能力.

2.5 试题拟合程度分析

表 2 统计了 8 道试题的拟合指标统计情况,包括试题难度、未加权残差均方(Outfit MNSQ)、加权残差均方(Infit MNSQ)、以及 PT 测量中的相关系数.

表 2 试题的拟合情况统计表

题号	试题难度	Outfit MNSQ	Infit MNSQ	相关系数
1	-3.49	1.07	4.06	0.11
2	-1.07	0.97	0.85	0.37
3	-1.60	0.93	0.64	0.37
4	0.98	0.97	0.98	0.49
5	0.59	0.99	0.98	0.46
6	0.08	1.11	1.17	0.35
7	1.98	0.94	0.90	0.53
8	2.53	1.02	1.03	0.47

根据拟合的数据来看,难度最大的题目为第 8 题,难度最小的题目为第 1 题.试题的加权残差均方全都在[0.93,1.11]范围内,未加权残差均方大部分都在[0.64,1.17]范围内,几乎都在正常的取值范围[0.5,1.5]内,表明试题的拟合程度较好.此外,相关系数表示试题与试题测量目标的拟合程度,相关系数的最低可接受水平约为 0.03,相关系数越高,说明试题与试题的测量目标越接近^[4].从表格中发现

所有试题的相关系数均处于可接受的水平.

2.6 试题气泡图分析

图 5 为试题的气泡图,直观地展示每一道试题的测量误差和拟合情况.图中,纵轴表示试题难度,越靠近顶端,表示题目越难;横轴表示试题与模型的拟合情况,用未加权残差均方表示;每个气泡的大小表示标准误,气泡越大,标准误越大,测量的精确性越低.

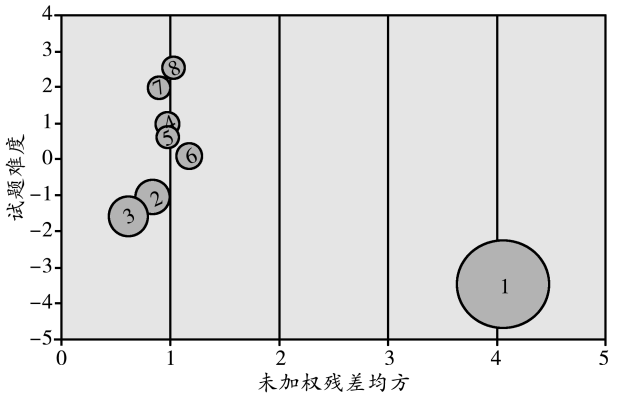


图 5 气泡图

由图 5 可知,2~8 题的未加权残差均方在 [0.5,1.5] 范围内,结合表 2 的拟合程度统计表,表明大部分学生的反应与预期一致,试题能够较准确地检测出学生的能力水平,以第 4 题为例进行说明,并对超出理想范围的第 1 题做进一步分析.

【第 4 题】国庆 70 周年阅兵展出了我国高超音速乘波体导弹——东风-17,东风-17 突防能力强,难以拦截,是维护祖国和平发展的有力武器.如图 6 所示,设弹道上处于大气层外的 a 点和处于大气层内的 b 点的曲率半径之比为 2:1,导弹在 a 和 b 两点的速度大小分别为 3 倍音速和 12 倍音速,方向均平行于其正下方的水平地面,导弹在 a 点所受重力为 G ,在 b 点受到空气的升力为 F .则()

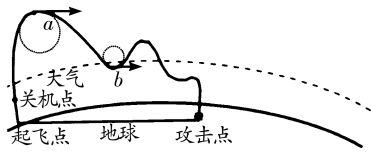


图 6 第 4 题题图

- A. $F=33G$
- B. $F>33G$
- C. $F=32G$
- D. $F<32G$

第 4 题的未加权残差均方为 0.97,在理想范围内,表明学生的作答情况符合模型的预期,即在本试题中,理解分析能力高的学生回答正确的概率大于

能力低的学生. 本试题以真实情境为背景, 利用向心力知识解决实际问题. 假设导弹在 a 和 b 点做圆周运动, 可以计算出升力 $F=33G$, 又从轨迹中得知导弹要做离心运动, 可得 $F>33G$. 62% 的学生答对了此题, 其中 42% 的学生能力水平值高于平均能力水平值, 能力高的学生回答正确的概率大于能力低的学生.

【第 1 题】某质点从静止开始做匀加速直线运动, 已知第 3 s 内通过的位移是 x , 则质点运动的加速度为()

- A. $\frac{3x}{2}$ B. $\frac{2x}{3}$
C. $\frac{2x}{5}$ D. $\frac{5x}{2}$

通过 Rasch 分析, 第 1 题难度为 -3.49 logit, 有 98% 的学生回答正确, 是客观题中最简单的题目. 本试题只考查了匀变速直线运动公式的运用, 学生无论能力高低均能正确回答, 学生的作答差异不大, 没有体现出必要的区分性.

本试题应该创设真实的问题情境, 让学生应用匀变速直线运动的物理知识解决生活中的实际问题. 这样才能考查学生是否真正理解所学知识, 同时也培养了学生分析问题、解决问题的能力, 有效发展学生的物理学科核心素养.

综上所述, 试题整体信度较高、区分度合理, 能客观地考查出学生分析问题的能力和水平. 但是相对来说, 缺乏考查高能力水平的试题, 存在个别试题指标与 Rasch 模型不能很好拟合的情况, 需要考虑做出调整.

3 Rasch 模型在试卷质量分析中的注意事项

Rasch 模型具有许多分析功能, 由于实际的测量情况不同, 测量者应该结合测量目标, 选择合理的 Rasch 模型对应的分析功能对测试进行分析. 本文以一份高三物理试卷中客观题为例, 主要从整体质量检验、怀特图、单维性检验、试题拟合度以及试题气泡图等方面进行质量分析. 如若教师想分析男、女生在作答试题上是否存在差异, 可以使用 Rasch 模型的项目功能差异 (Differential Item Functioning, DIF) 进行检验; 对于考查多种潜在特质的试题, 教

师可以使用多维度的 Rasch 模型进行分析. 多维度的 Rasch 模型并不是对 Rasch 模型的颠覆, 而是对 Rasch 模型单维度要求的发展, 它充分利用相关维度特质 (或相关分量表) 所提供的有用信息, 在保证测验信度、效度的同时, 提高目标特质测量的精确度、广度和效率^[1].

在分析过程中, 如果有试题的拟合程度不理想, 我们不应该直接删除该试题而是要认真地审视该试题找出影响拟合度的因素, 如考虑命题人员的初始意图、测评目标与要求等, 再决定这类试题的去留. 因为 Rasch 模型作为检测工具, 只能检测出有问题的试题, 对于如何处理这类试题, 需要结合实际情况做出判断.

4 结论

本文以一份高三物理试卷中客观题的质量分析为例, 介绍了 Rasch 模型在物理学业质量评价中的应用. 研究表明试题整体质量较好, 信度较高, 区分度合理, 绝大多数试题达到了测量目标, 能客观地考查学生的能力水平. 可以看出, 中学教师编制该试题的能力水平较高, 能够客观、准确地评价学生的物理学业质量.

Rasch 模型的分析功能比较多, 本文中所提及的功能只是 Rasch 模型可用于试卷质量分析功能的一部分. 因此在具体的应用中, 由于实际的测量情况不同, 测量者应该结合测量目标, 选择合理的 Rasch 模型对应的分析功能对测试进行分析^[5].

参考文献

- 1 晏子. 心理科学领域内的客观测量[J]. 心理科学进展, 2010(8):1298~1305
- 2 Maja P, Lana I, Ana S. Rasch model based analysis of the Force Concept Inventory[J]. Physical Review Special Topics Physics Education Research, 2010(1):165~186
- 3 Bond T G, Fox C M. Applying the Rasch model: Fundamental measurement in the human sciences[M]. Psychology Press, 2006
- 4 赵守盈. Rasch 模型在研究生入学考试质量分析中的应用[J]. 教育研究, 2012(6):61~65
- 5 柏毅, 朱文琴, 陈慧珍. Rasch 模型在试卷质量分析中的应用——以小学科学六年级技术与工程素养测试卷为例[J]. 教育测量与评价, 2019, 216(1):27~33