

基于Rasch模型的高中生物试题解析

李红彦, 杨雨怡

贵阳学院材料科学与工程学院, 贵州 贵阳

收稿日期: 2025年3月12日; 录用日期: 2025年4月11日; 发布日期: 2025年4月22日

摘 要

Rasch模型作为一种现代教育测量工具, 以其精准的反馈在试题分析和学生能力评估中发挥重要作用。本研究以高中生物测验为例, 从单维性检验、试卷拟合度、题目拟合度、怀特图和气泡图等方面展开分析, 采取数据与试题内容相结合的方式对试题进行解析, 依据数据与试题解析结果, 提出试题分析需将所得数据与试题内容相结合、合理设置题目难度系数, 提高试卷结构的科学性和区分度、创新试题形式, 加强生活化情境设计等建议。

关键词

Rasch模型, 高中生物, 试题分析

Analysis of High School Biology Test Items Based on the Rasch Model

Hongyan Li, Yuyi Yang

College of Materials Science and Engineering, Guiyang University, Guiyang Guizhou

Received: Mar. 12th, 2025; accepted: Apr. 11th, 2025; published: Apr. 22nd, 2025

Abstract

Rasch model, as a modern educational measurement tool, plays an important role in test analysis and student ability assessment with its accurate feedback. Taking high school biology test as an example, this study analyses the test questions in terms of unidimensionality test, test paper fit, items fit, Wright map and bubble chart, and analyses the test items by combining the data with the contents of the test items. Based on the results of the data and the analysis of the items, it is proposed that test items should be analysed by combining the data with the contents of the test questions, reasonably setting the question difficulty coefficients, and improving the scientificity and differentiation of the structure of the test paper; innovative forms of test items and strengthening the design of life situations.

Keywords

Rasch Model, High School Biology, Test Analysis

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在新高考、新课程和新教材“三新”背景下,传统的高中生物教学方式难以满足当前教学需求。新课程和新高考更加强调培养学生的核心素养,以提升学生的综合素质。由于传统的考试和评价方法难以全面反映学生的生物学素养和能力。因此,引入新的评估模型尤为重要。Rasch 模型作为一种项目反应理论(IRT)模型,在心理学、社会学等领域得到了广泛应用。该模型可量化学生能力和试题难度,为全面评估学生核心素养提供了条件。然而,在高中生物教学中,基于 Rasch 模型对试题解析和学生能力评估的研究相对较少。将 Rasch 模型应用于高中生物试题的解析中,不仅能够更准确地评估学生能力、试题难度和设计有效性,还能为教学及试题改革提供科学依据。这对于提高教学质量、优化教学内容、深化试题改革具有重要意义。本研究基于 Rasch 模型对高中生物试题进行解析,按照测量指标的选择、数据与试题分析、结果解释、结论与建议这一思路依次进行,通过分析数据及试题内容,为高中生物教学提供更精准的反馈。

2. Rasch 模型简介

Rasch 模型由丹麦数学家和心理学家于 1960 年提出,是一种用于心理测量和教育评估的统计模型,主要用于分析个体在测试中的反应数据,评估被试的能力与题目难度之间的关系,属于项目反应理论(IRT)的范畴。Rasch 的简单函数表示为: $\text{Loge}(\text{Pni}/(1 - \text{Pni})) = \text{Bn} - \text{Di}$ [1], 其中, Pni 表示被试者 n 对项目 i 回答正确的概率, Bn 表示被试者 n 的能力参数, Di 表示项目 i 的难度参数。目前, Rasch 模型适用教育测量和心理测量等领域,用于项目筛选、能力评估、测试设计和信度效度分析等,在教育领域的研究聚焦于学生能力、测试评分研究、试卷分析、教师能力、课程规划等[2]。此外, Rasch 模型具有样本独立性、项目独立性和测量不变性等优势,该模型通过被试对试卷中试题的反应来测量其潜在特质。

3. 研究方法

3.1. 高中生物测评试卷

本研究使用高中三年级生物测评试卷,共计 32 题,其中客观题(选择题、填空题) 16 题,主观题 16 题。客观题以 O01~O12、M13~M16 表示,主观题以 S1701~S2103 表示,其中, S1701 表示第 17 题第 1 问,以此类推。

3.2. 样本选择

以某高中三年级全体学生为样本,本次测试,共获得有效数据 393 人。

3.3. 分析方法

本次数据分析主要采用了 Excel 和 Winsteps 5.4.3 两个软件,Excel 用于处理原始数据,处理完毕后,

将数据导入 Winsteps 5.4.3 软件中进行进一步分析。Winsteps 是专门用于 Rasch 模型数据分析的软件, 其在教育测验领域得到了广泛应用。本次数据分析, 选择了分部计分模型(Partial Credit Model, PCM)进行数据处理。分析内容主要包括: 单维性检验、试卷拟合度、题目拟合度、怀特图(Wright Map)、气泡图(Bubble Diagram)等。

4. 数据分析结果

4.1. 单维性检验分析

单维性检验是项目反应理论(IRT)中的一个重要假设检验, 它指的是测验旨在测量被试的某一种能力, 可以忽略其他能力对测验结果的影响。单维性检验分析主要用于评价试题是否紧扣学科主题。若试题展现出良好的单维性特征, 说明学生在考试过程中主要受生物能力的潜在特质影响, 从而影响作答表现。

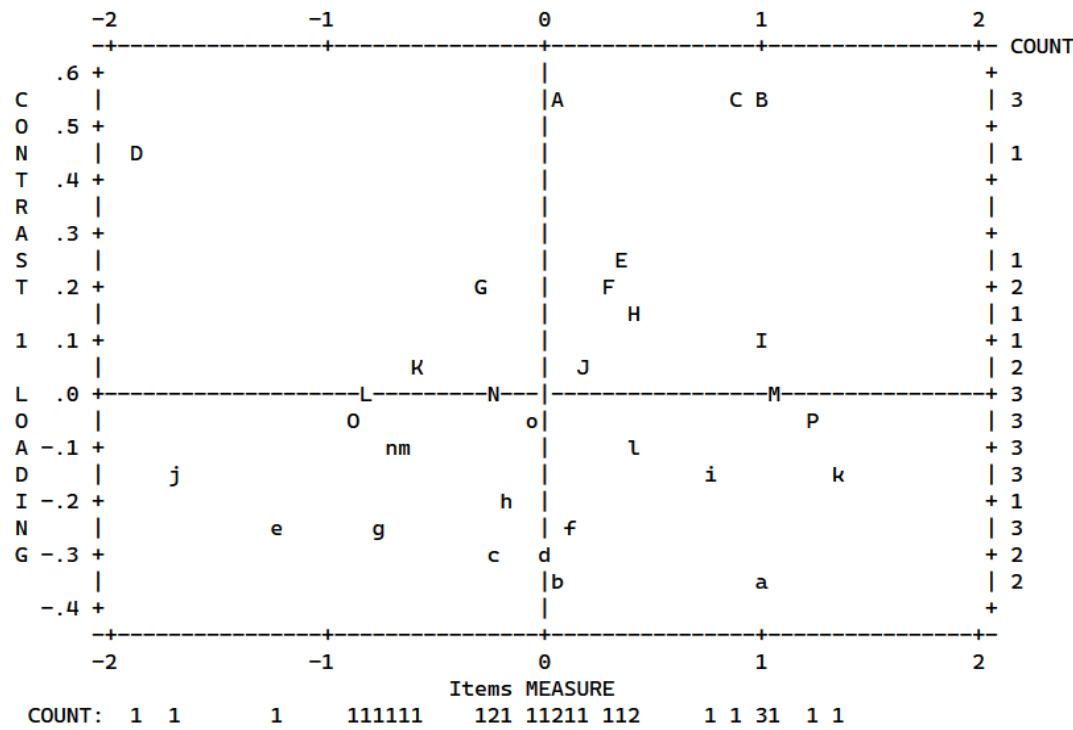


Figure 1. One-dimensionality test
图 1. 单维性检验

本次测试的单维性检验结果如图 1 所示, 横轴表示项目的测量值(Items Measure), 范围从-2 到 2。纵轴表示不同的题目(Items), 从 A 到 K。根据图 1 的单维性检验结果, 可知大部分题目的项目测量值(Items Measure)集中在[-0.4, +0.4]区间内。这一分布符合 Rasch 模型理论中对单维性的要求, 意味着这些题目主要受到单一潜在因素的影响。图 1 显示的题目在潜在变量上的分布较为集中, 进一步支持了这些题目具有单维性, 即它们共同测量了一个共同的潜在特质或因素。但 A、B、C、D 代表的 4 个题目的相关关系值超出[-0.4, +0.4]的建议范围, 不符合单维性检验要求, 表明 S2001、S2002、S2003、M14 这 4 道题目不只受单一因素影响, 可能存在其它潜在特质影响了学生作答。

图 2 是基于方差的单维性检验结果, 图中的“1”代表由题目解释的方差, 而“1”代表未解释方差主成分分析中第 1 主成分。若 $I > 1$, 则说明题目解释的方差高于残差中第一主成分解释的方差。换言之,

即使数据中存在第二个能力维度对测量结果产生一定影响，也不影响测量结果的单维性。图 2 数据表明，I 为 19.5%，而 1 为 4.1%，I 明显高于 1。这说明尽管 A 和 a 可能受到多个潜在特质的影响，但并未影响整体测试结果。总之，该试题满足 Rasch 模型的单维性要求。因此，可以进行后续的分析。

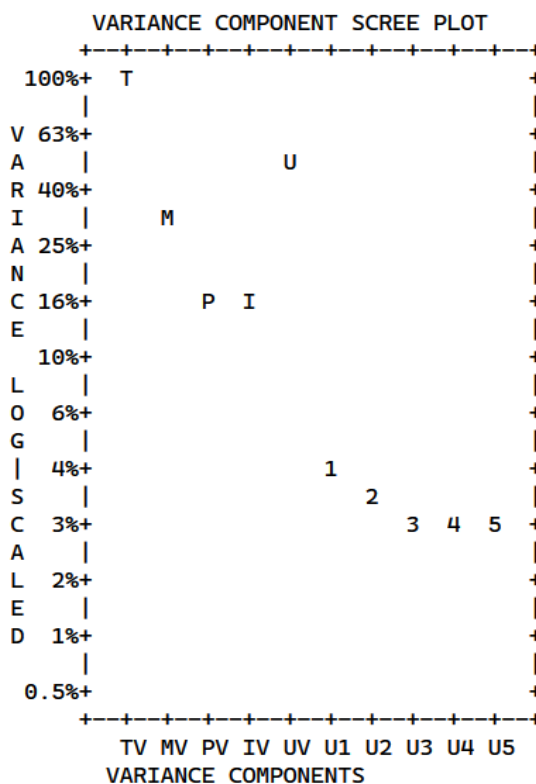


Figure 2. Principal component analysis
图 2. 主成分分析

4.2. 试卷拟合

对 393 名学生的数据进行整体质量检验分析，结果如表 1 所示。

Table 1. Overall quality test results

表 1. 整体质量检测结果

	数量	INFIT		OUTFIT		Speration	Reliability
		MNSQ	ZSTD	MNSQ	ZSTD		
被试(Student)	393	1.00	0.0	1.03	0.1	1.67	0.75
项目(Item)	32	1.00	0.1	1.05	0.2	8.25	0.99

表 1 列出了两组数据：被试(Student)和项目(Item)，每组数据都包括数量、Infit 和 Oufit 的 MNSQ (均方差)和 ZSTD (标准差)，以及区分度(Separation)和信度(Reliability)，以上数据用来检测试卷的整体质量。

ZSTD 是 MNSQ 的标准化形式，表示拟合统计量的显著性。当 IMNSQ 的理想值为 1 时，表示实际数据与 Rasch 模型完全拟合；当 IMNSQ 的值在[0.5, -1.5]之间时，表示数据与模型预期拟合程度可接受；理想拟合状态下 ZSTD 值为 0；当 ZSTD 值在[-2, -2]时，认为拟合较[3]。由表 1 可知，对于被试(学生)

组，其 IMNSQ 值为 1.00，ZSTD 值为 0.0，表明数据的拟合度较好。OMNSQ 值为 1.03，ZSTD 值为 0.1，也显示出良好的拟合度。一般项目与被试的区分度均需大于 2 [4]，信度理想值为 1，大于 0.7 表明测验信度较高，0.6~0.7 表明信度良好[5]。本次测试区分度为 1.67 (<2)，信度为 0.75 (>0.7)，说明数据的区分度较低，可靠性较高。对于项目(题目)组，其 IMNSQ 值为 1.00，ZSTD 值为 0.1，显示出良好的拟合度。OMNSQ 值为 1.05，ZSTD 值为 0.2，也表明拟合度较好。分离度为 8.25.31，信度为 0.99，说明项目的区分度和可靠性非常高。总体来看，被试的数据除了区分度未在可接受范围内，其拟合度和可靠性均在可接受范围内，项目的数据质量都较好，拟合度、区分度和可靠性均在可接受的范围内。

4.3. 题目拟合

本次测验，试卷的 32 个题目拟合统计信息如表 2 所示。

Table 2. Item fit statistics table
表 2. 题目拟合统计信息表

题目	难度	标准误	Infit MNSQ	Outfit MNSQ	相关系数
O01	-0.720	0.121	1.01	0.99	0.20
O02	-0.207	0.110	0.91	0.88	0.41
O03	-0.592	0.118	1.04	0.07	0.12
O04	1.056	0.107	0.97	0.98	0.27
O05	1.022	0.106	0.97	0.98	0.27
O06	1.225	0.140	0.92	0.87	0.36
O07	0.432	0.104	1.06	0.07	0.11
O08	1.679	0.164	0.98	0.91	0.22
O09	-0.255	0.110	1.00	0.00	0.22
O10	-0.905	0.127	0.97	0.98	0.26
O11	-0.634	0.119	1.00	0.05	0.19
O12	-0.795	0.124	1.00	0.99	0.20
M14	-1.888	0.229	1.06	1.91	0.01
M15	0.328	0.061	1.33	1.53	0.05
M16	-0.036	0.065	1.20	1.39	0.17
S1701	-0.762	0.064	0.90	0.89	0.47
S1702	0.187	0.039	1.06	1.08	0.49
S1703	0.316	0.086	1.05	1.27	0.16
S1801	-0.161	0.046	0.90	0.89	0.55
S1802	0.096	0.055	0.91	0.91	0.50
S1803	0.421	0.047	1.00	0.99	0.47
S1901	-0.317	0.081	1.14	1.15	0.07
S1902	0.041	0.075	1.00	1.04	0.35
S1903	0.014	0.053	0.87	0.84	0.55
S2001	1.003	0.060	1.06	1.08	0.30

续表

S2002	0.039	0.060	0.94	0.89	0.45
S2003	0.887	0.042	1.02	1.02	0.48
S2004	1.206	0.066	0.89	0.88	0.48
S2101	0.782	0.039	1.02	1.02	0.53
S2102	1.352	0.056	0.94	0.94	0.41
S2103	0.996	0.046	0.95	0.96	0.51

表 2 显示了评测试卷每道题目的拟合指数统计,从表中可以看出,不同题目的难度系数、标准误差、Infit MNSQ、Outfit MNSQ 和相关系数各不相同。这些指标可以帮助评估每个题目的质量和拟合度,从而为测试的可靠性和有效性提供依据。有研究者建议, Infit MNSQ 和 Outfit MNSQ 的取值应当在[0.8, 1.2]之间,较为宽松的标准认为取值应当在[0.5, 1.5]之间[6]。作为高风险考试,应当选择较为严格的拟合标准[7]。从表 2 可知,各题目 Infit MNSQ 的取值范围在[0.87, 1.33]之间,大多数题目都在可接受的范围内,表明数据与模型拟合较好。但 M15 参数值为 1.33 (>1.2),超过阈值,说明该题与模型的拟合度较差。因此,需要对 M15 题的试题内容进行分析,了解该题与模型拟合度较差的原因。M15 题如图 3 所示,这道题不仅考查控制变量法和植物激素知识点,还考查学生对探究实验的理解和数据分析能力,研究者探究了不同浓度和作用时间下 6-BA 对拟南芥种子萌发和幼苗生长的影响,要求学生分析表格数据,选出正确选项,选项涉及实验自变量、细胞分裂素的合成部位、6-BA 与赤霉素的功能比较,以及特定浓度 6-BA 的作用效果。

15. 6-BA 是一种细胞分裂素类似物,研究者探究了不同浓度和作用时间下 6-BA 对拟南芥种子萌发和幼苗生长的影响,结果如下表。下列分析正确的是

培养天数	处理	0.2mg/L	0.4mg/L	0.8mg/L	1mg/L	1mg/L	2mg/L	4mg/L	8mg/L	CK
		18h	18h	18h	18h	7h	7h	7h	7h	18h
13 天	发芽率	60%	80%	100%	60%	60%	60%	80%	40%	72%
	每株叶片数	6	7	7	5	4	5	3	2	5
21 天	长势	a ⁺ b ⁺ c ⁺	a ⁺⁺ b ⁺⁺ c ⁺	a ⁺⁺ b ⁺⁺ c ⁺⁺	a b c	a ⁺⁺ b c	a ⁻ b ⁻ c ⁻	a ⁺⁺ c ⁺	植株畸形	a b c

注:CK 为空白对照组;a:叶片大小,b:叶色,c:植株大小;+:高于对照;-:低于对照。

- 本实验的自变量有 6-BA 的浓度和处理时间
- 植物体内细胞分裂素的主要合成部位是根尖和嫩叶等
- 在促进种子萌发方面,一定浓度的 6-BA 与适宜浓度赤霉素具有相似功能
- 1mg/L 6-BA 对拟南芥的发芽率、每株叶片数及长势有一定的抑制作用

Figure 3. Item 15

图 3. 第 15 题

经分析,该题不符合模型预期表现的原因是题目提供的数据表信息量大,学生在解答此题时,需要在短时间内分析多个指标、区分自变量和因变量以及它们之间的关系,还需要运用细胞分裂素、赤霉素等相关储备知识进行迁移。此外,题目要求学生具备较强的探究实验理解和数据分析能力。然而,部分学生难以区分探究实验中的自变量和因变量,从而影响他们对每个选项正误的准确判断。对此,需将复杂的数据表简化,突出关键信息,减少学生的信息处理负担,还可以设计一些具体的情景或案例,让学

生在实际问题中应用所学知识,这有助于提高题目的区分度和拟合度。据数据显示, Outfit MNSQ 的取值范围在[0.00, 1.91]之间,其中 O03、O07、O09、M14、M15、M16 和 S1703 参数值分别为 0.07、0.07、0.00、0.05、1.91、1.53、1.39 和 1.27,均偏离了正常范围。除 M15 题外其它六道题的 Infit MNSQ 的取值范围在[0.80, 1.20]范围内,综合两个拟合参数,可以认为 O03、O07、O09、M14、M16 和 S1703 六个题目在可接受范围内,M15 题确实受到其它因素干扰。此外,M13 题出现天花板效应,无法进行参数估计。

4.4. 怀特图

怀特图(Wright map)是一种用于展示项目反应理论(Item Response Theory, IRT)中项目和受试者之间关系的图表。它通常用于心理测量学和教育测量学中,以分析测试题目和学生的能力。本次测试的题目难度与学生能力之间的关系,如图 4 所示。

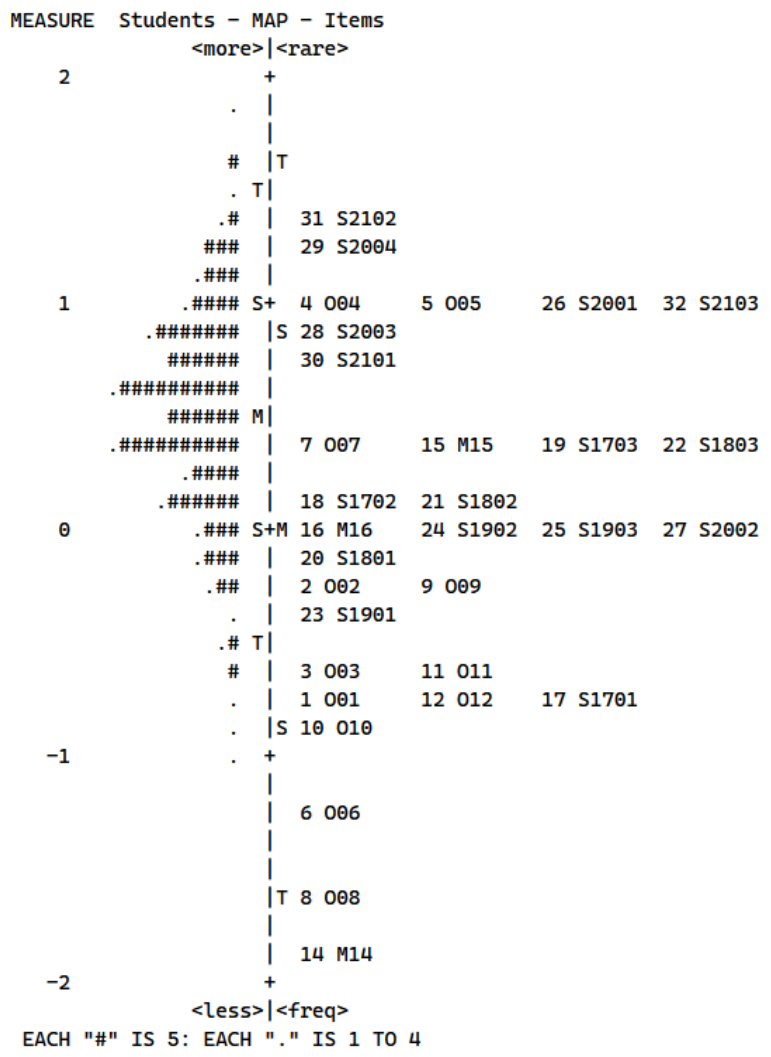


Figure 4. Wright map
图 4. 怀特图

在怀特图(图 4)中,中间的竖线是 Logit 刻度尺,横轴和纵轴表示题目的难度和考生的能力。#表示题目的难度等级,每个#代表一个单位的难度增加。S+和 S 表示受试者的能力等级,S+表示高于平均水平的

能力, S 表示平均或以下的能力。M 表示中等难度的题目, T 表示测试题目。题目和学生之间的距离表示它们在特征值上的接近程度。例如, 一个 S+ 学生和一个 M 难度的题目之间的距离较小, 表示这个学生可能能够正确回答这个题目。

从 Logit 标尺左侧可以看出, 被试学生的能力水平处于 $[-0.2, 1.2]$ 之间, 能力分布范围为 1.4Logit, 平均值为 0.511Logit。但在 Logit 值大于 1.3 时, 没有与学生能力水平相适应的题目, 表明该试题缺少难度大的题目。与之相对应, 题目难度平均值被设定为 0, 二者之差为 0.511Logit, 说明该试题对于被试来说整体难度偏低。

Logit 标尺右侧为题目, 从分布位置上看, 题目 S2102 难度最大, 该题属于认知层次中的理解维度, 考察学生的科学探究素养和理性思维素养, 其难度为 1.352Logit。M14 难度最小, 因为该题属于基础知识, 考察学生学科核心素养中的生命观念, 其难度为-1.888Logit, 题目难度的分布范围为 3.24Logit。

学生能力均值较试题难度均值高 0.511Logit, 由此可以看出学生的能力水平相对高于测验项目的难度水平。这说明项目难度设计与学生的实际水平不匹配, 项目难度偏低。而第 6、8、14 题过于简单, 因为这三道题目均考察学生对基础知识概念的理解, 第 6 题考察学生的社会责任素养、第 8 题和第 14 题考察学生的生命观念素养。所以, 学生对于这种题型容易答对。由此可知, 本次考试, 存在一定数量的简单题。与简单题目较多的情况相对应, 测验中的难题较少, 尤其是在 Logit 值超过 1.3 时, 没有相应的难题与学生能力匹配。总体看, 试题中简单题目较多, 试题难度差异较大, 中等难度试题较为集中, 不利于区分对不同能力水平的学生。

4.5. 气泡图

气泡图用于展示不同题目(Items)在两个维度上的分布情况。本次数据生成的气泡图如图 5 所示。

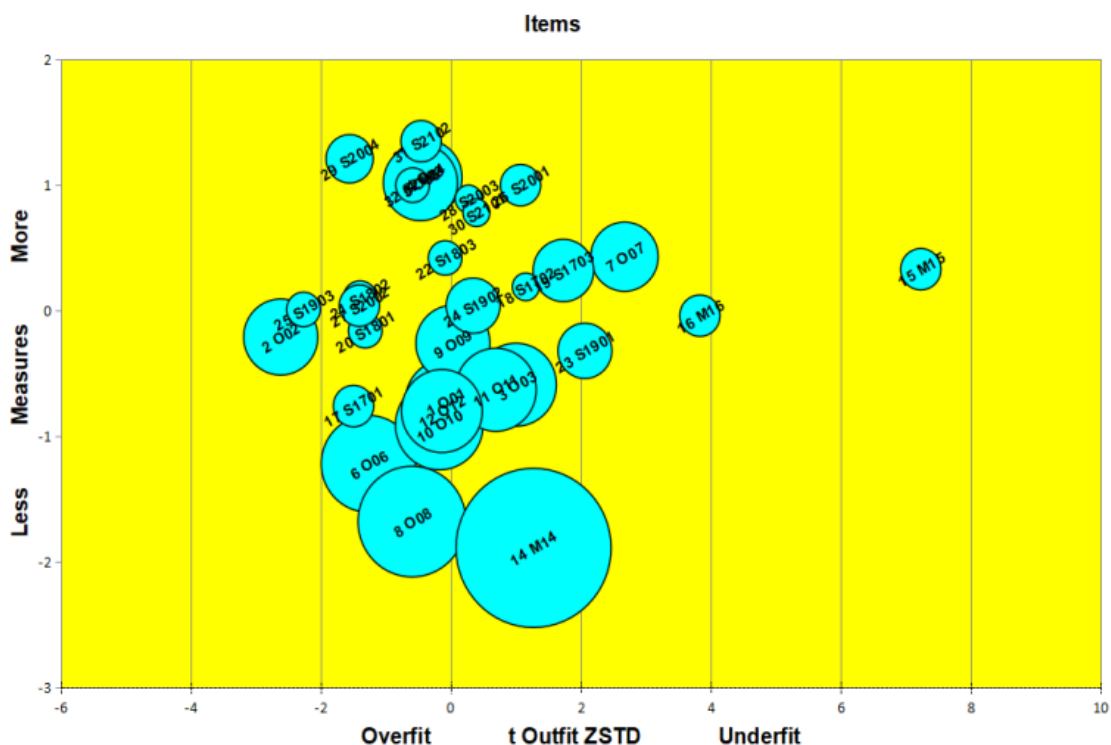


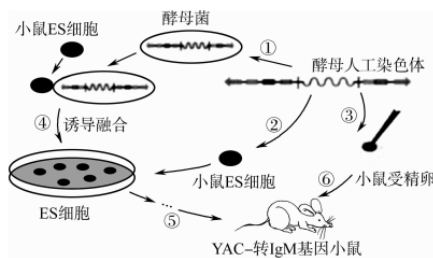
Figure 5. Bubble diagram

图 5. 气泡图

在图 5 中, 横轴表示“Overfit”和“Underfit”之间的差异, 纵轴表示“Less”和“More”之间的差异。每个气泡代表一个题目, 气泡的大小代表 Rasch 模型标准误, 气泡越小说明误差越小, 测量结果越精确; 气泡左右位置代表项目拟合度, 靠左表示过渡拟合、靠右代表不拟合, 气泡越靠近气泡图的中轴线则说明题目与模型拟合得越好[8]; 气泡越靠近上方, 说明题目的难度越大[9]。

该气泡图展示了不同题目在“Overfit”和“Underfit”以及“Less”和“More”两个维度上的分布情况。大多数题目集中在左下角, 落在了可接受区间 $[-2, 2]$ 内, 表示它们在这两个维度上都有较低的值, O02、S1903 为过拟合, 代表这些项目与模型拟合度高。而一个较大的气泡位于右下角, 表示该题在这两个维度上都有较高的值; M15、M16 不在可接受区间内, 再欠拟合(underfit)区间内, 表明有能力高的做错而能力低的做对此题的情况[10]; 因此, 也需要对 M16 的试题内容进行分析, 了解能力高的做错而能力低的做对此题的原因, 以修正题目。M16 题如图 6 所示, 主要考查学生对酵母人工染色体(YAC)构建及相关生物技术的理解 and 应用能力, 该题涉及基因工程、细胞工程和胚胎工程等多个知识点。题目给出了利用整合有人源抗体 IgM 基因的 YAC 培育能分泌人源抗体小鼠的三个技术路线图, 要求学生判断哪个叙述是错误的。

16. 酵母人工染色体(YAC)是由酵母细胞中必需的复制起始序列等与目的基因和质粒构建成的表达载体, 具有易在真核细胞中表达的特点。下图表示利用整合有人源抗体 IgM 基因的 YAC, 培育能分泌人源抗体小鼠的三个技术路线图。下列叙述错误的是



- A. 构建 YAC 时用到的工具酶有 DNA 聚合酶、DNA 连接酶和限制酶
- B. 图中将 YAC 导入受体细胞时均需用钙离子处理受体细胞
- C. 从 YAC—转 IgM 基因小鼠体内提取到的抗体即为所需单克隆抗体
- D. 培育 YAC—转 IgM 基因小鼠的三个技术路线中均需用到胚胎移植

Figure 6. Item 16

图 6. 第 16 题

经分析, 该题不符合模型预期表现的原因有以下几点: 一是知识点综合性强, 题目涉及 YAC 的构建、基因工程、细胞工程和胚胎工程等多个知识点, 学生需要理解并综合运用这些知识才能准确解答; 二是学生认知水平与题目要求的匹配度欠佳, 题目要求学生具备较高的综合分析和应用能力, 对于能力高的学生来说, 可能超出了他们的认知水平, 导致他们选错。因此, 可将题目拆分为更基础的知识点, 降低综合难度, 或者在题目中增加生活化情境, 帮助学生更好地理解题目要求, 如在题目后补充: 人源抗体 IgM 基因的 YAC 载体就像一个“基因运输车”, 它能将人类抗体 IgM 基因安全地运送到小鼠细胞中, 并确保这些基因能够在小鼠体内正确表达。

M14、O08 和 O06 气泡较大, 表明其难度估计的误差较大, 其对应测量结果的精准性较小。S2102 和 S2004 气泡位置在顶端, 说明这两道题难度最大。综合 Rasch 模型分析的数据来看, 第 14 题考查的内容是生命观念, 该题主要考查基因、DNA 和性状的关系, 要求学生识记基因对性状控制的方式。学生识记该知识点后, 便能轻松选出答案。该题难度为 -1.888 , 难度最低, 其内部拟合值 M 为 1.06 , 外部拟合值

M 为 1.91, 可以看出, 该题目外部拟合值 M 值较高并超出了理想范围。在 Rasch 模型中, M 值用来评估项目难度水平。当外部拟合 M 值过高时, 表示该题目的难度被高估。综上, 第 14 题的难度过低, 且主要考查学生对知识点的识记, 并没有设计相应情景来引导学生思考。

第 S2102 题以科学探究题目的形式综合考查学生对知识的理解及灵活应用知识的能力, 以分离能有效分解纤维素的菌株为背景, 考查微生物的分离和培养, 要求学生识记纤维素酶的概念、灭菌和消毒的区别等知识, 所以难度较大。该题目难度为 1.352, 是试卷所有题目中难度最高的题目, 对于难度较大的题目可能会影响学生能力区分度, 因为题目难度大时, 学生之间的得分差异较小, 从而难以区分学生个体之间的能力水平。

5. 结论与建议

5.1. 测试评价

经检验得出, 该生物试题大部分题目符合单维性要求, 能够较好地反映出学生的生物学科能力, 但有 4 道题目的单维性检验结果较差: S2001、S2002、S2003、M14, 说明学生在作答这 4 道题目时, 受到了除生物能力之外的潜在特质的影响。此外, 该试题具有良好的信度和区分度, 能够准确测量并区分学生的能力水平。然而, 整体试题难度偏低, 缺乏高难度题目, 尤其是在 Logit 值超过 1.3 时, 缺乏与学生能力水平相匹配的题目, 因此需要对试题难度进行调整, 并增加一些高难度的题目。最后, 部分题目需要进行优化, 如 M13 题出现了天花板效应, 导致无法进行参数估计。

5.2. 相关建议

1) 试题分析需将所得数据与试题内容相结合

在使用 Rasch 模型分析高中生物试题时, 应将数据与试题内容紧密关联, 以便更准确地了解数据所揭示的问题根源。例如, 对于单维性检验结果较差的题目, 应深入分析其知识点覆盖范围和题目表述, 看是否涉及多个知识点的考查或题目表述不清, 导致学生理解困难。如图 7 第 20 题的第(3)小题, 题目分为两小问且表述较为复杂。

20. (12分)果蝇生活史短、易饲养、繁殖快, 是很好的遗传学实验材料。果蝇的体色有灰体和黑体, 由等位基因(M/m)控制, 眼色有红眼和棕眼, 由等位基因(N/N)控制。已知两对基因独立遗传, 与眼色相关的某种基因型存在致死现象。为探究上述两对性状的遗传规律, 某小组进行了相关的杂交实验:
- 实验一: 灰体红眼 早 × 黑体棕眼 ♀ → 灰体红眼 早 : 黑体红眼 早 = 1: 1
- 实验二: 黑体红眼 早 × 灰体棕眼 ♀ → 灰体红眼 早 : 灰体棕眼 早 : 灰体棕眼 ♀ = 1: 1: 1
- 回答下列问题:
- (1) 果蝇体色性状的遗传方式为 _____, 判断的依据是 _____。
- (2) 实验一的 F₁ 均为雌果蝇, 原因是 _____。
- (3) 实验一的 F₁ 雌果蝇的基因型是 _____。将实验一中的 F₁ 雌果蝇与实验二中的 F₁ 雄果蝇随机交配, 子代雄果蝇的表型及比例为 _____。

Figure 7. Item 20

图 7. 第 20 题

第一小问要求写出实验一的 F₁ 雌果蝇的基因型, 学生需根据实验一的杂交组合和结果, 结合体色与眼色基因的独立遗传及致死现象, 推断出 F₁ 雌果蝇的基因型。第二小问要求计算将实验一中的 F₁ 雌果蝇

与实验二中的 F_1 雄果蝇随机交配后, 子代雄果蝇的表型及比例, 这需要学生对两对基因的组合情况、交配过程以及致死基因型的影响进行综合分析, 对学生的逻辑思维和计算能力要求较高。

因此, 可将题目拆分为两个独立的小问题, 引导学生分步骤思考, 降低理解难度。基于 Rasch 模型的试卷分析结果不是独立于应用存在的, 其结论应结合评价的目的、命题设计等要素进行综合考量。在实际应用工作中, 需要警惕仅关注拟合指标的唯参数论。

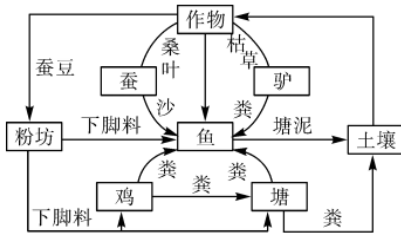
2) 合理设置题目难度系数, 提高试卷结构的科学性和区分度

新一轮的课标修订中指出要科学合理地设置测评试卷的难度系数, 防止偏题、难题、怪题的出现, 试卷题目难度系数比例要合理, 试卷总体结构要规范、要科学并且能够通过测评试卷有效区分学生能力水平。题目难度过低, 可能无法有效区分不同能力水平的学生, 以至于高分学生和低分学生之间的差距不明显, 从而影响考试的公平性和选拔功能; 题目难度过高, 可能会导致大多数学生都无法正确回答, 从而无法区分考生之间的能力差异, 也难以准确评估学生在特定领域的知识水平。因此, 设计测试题目时, 需要确保题目难度适中, 既能考察考生的基本能力, 又能区分不同水平的考生, 以实现测试的目的和功能。具体操作上, 可将试卷结构分为基础题(占 40%~50%)、中等难度题(占 40%~50%)和高难度题(占 10%~20%), 分别对应考查核心概念、知识理解和综合应用能力。例如, 在生物试卷中, 基础题可考查细胞结构等基本概念, 中等难度题要求学生分析实验数据得出结论, 高难度题则要求学生结合复杂生态情境, 综合运用知识解决问题, 以实现测试的选拔与区分功能。

3) 创新试题形式, 加强生活化情境设计

测评试题的设计应融入生活化情境测评, 形成具有生活化的试题情境, 如加强试题情境与学生生活、社会实践的联系, 注重考查学生在具体情境中综合运用所学知识分析问题和解决问题的能力。创设试题情境不仅能提升学生的作答兴趣, 还能简单明了地向学生传达题目的意思, 防止学生因误解题目而影响作答效果。此外, 生活化情境试题还能够使学生“身临其境”地理解问题、思考问题并作答。在设计试题时, 应注重将学科知识与学生熟悉的生活场景相结合, 使试题更具情境性, 从而提高学生的理解能力。例如, 图 8 第 19 题仅呈现了某农业生态示意图, 要求学生看图理清流程并完成相应题目, 题目缺乏情境性。

19. (10 分) 如图是某农业生态园示意图, 回答下列问题:



- (1) 作物属于该生态系统的 _____; 作物为该生态系统的主要成分, 理由是其可为其他生物提供 _____。
- (2) 塘泥、粪等通过土壤微生物的 _____ 能为作物的生长提供 _____ (填营养成分)。
- (3) 该生态农业中, 农田、粉坊、鸡舍、鱼塘、驴场、蚕室合理布局, 运行稳健, 主要体现了生态工程的原理有 _____ (答出两点即可)。农业生态园将农业活动、休闲娱乐等融为一体, 可供游客旅游观赏, 这体现了生物多样性的 _____ (填“直接”或“间接”) 价值。

Figure 8. Item 19
图 8. 第 19 题

该题可采用图文结合的形式呈现,在图的上方增加与题目相关的生活化情境信息,以帮助学生理解农业生态示意图。生活化情境设计:将农业生态园比作一个自给自足的“生态小镇”。作物是小镇的主要生产者,为居民(其他生物)提供食物和能量;塘泥、粪等废弃物通过土壤微生物的处理,就像小镇中的垃圾处理厂,将垃圾转化为有用的肥料,为作物的生长提供营养;整个生态园的合理布局和稳健运行,体现了生态工程的系统性,就像小镇的规划和管理,需要遵循生态平衡和可持续发展的原则。这种情境设计有助于学生理解生态系统的组成和生态工程的应用。

参考文献

- [1] 罗德红,龚婧. Rasch 模型在试卷质量分析中的应用——基于五六年级学生阅读素养前测试卷的质量分析[J]. 教育测量与评价(理论版), 2015(1): 18-22.
- [2] 梁晨,林世威. Rasch 模型在中国教育领域的应用研究[J]. 科技资讯, 2019, 17(35): 227-228, 230.
- [3] 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评测试卷为例[J]. 教育测量与评价, 2019(1): 25-31.
- [4] 关聪,崔雪梅. 基于 Rasch 模型的初中物理核心素养试题设计与质量分析[J]. 物理教师, 2024, 45(7): 28-32, 40.
- [5] 潘澎,胡象岭. 基于 Rasch 模型的力概念量表质量分析[J]. 物理教师, 2021, 42(8): 2-6.
- [6] 何壮,袁淑莉,赵守盈. 教育考试中短测验的分析方法——基于两种项目反应理论方法的比较研究[J]. 中国考试, 2012(10): 18-24.
- [7] 何壮,袁淑莉,余水,等. 心理测量在高风险考试分析中的应用[J]. 贵阳学院学报(社会科学版), 2020, 15(2): 114-118.
- [8] 徐柱柱. 基于 Rasch 模型的数学直观素养测验质量分析[J]. 理科考试研究, 2023, 30(16): 21-25.
- [9] 张迪. Rasch 模型在八年级数学学业测试卷质量分析中的应用[J]. 教育测量与评价, 2020(8): 34-41.
- [10] 高靖寒,柏毅. 基于 Rasch 模型的小学四年级科学素养评价研究[J]. 东南大学学报(哲学社会科学版), 2021, 23(S1): 135-138.