

基于Rasch模型的高中学生生物学科学建模能力测评初步研究

周宏伟, 王 威*

四川师范大学生命科学学院, 四川 成都

收稿日期: 2024年9月10日; 录用日期: 2024年10月9日; 发布日期: 2024年10月17日

摘 要

科学建模能力是科学思维能力重要组成部分, 其发展与评价应当成为教学工作的重点。本研究通过初步构建高中生生物学科学建模能力测评框架并尝试开发测评工具, 最终利用Rasch模型对工具进行检验, 进而证明工具的有效性, 为教师及其他研究者提供评价思路。

关键词

科学建模能力, Rasch模型, 评价研究

Preliminary Study on the Assessment of High School Students' Biological Scientific Modeling Competence Based on the Rasch Model

Hongwei Zhou, Wei Wang*

College of Life Science, Sichuan Normal University, Chengdu Sichuan

Received: Sep. 10th, 2024; accepted: Oct. 9th, 2024; published: Oct. 17th, 2024

Abstract

The ability to engage in scientific modeling is an important component of scientific thinking skills, and its development and assessment should become a focus of teaching work. This study preliminarily

*通讯作者。

文章引用: 周宏伟, 王威. 基于 Rasch 模型的高中学生生物学科学建模能力测评初步研究[J]. 教育进展, 2024, 14(10): 730-737. DOI: 10.12677/ae.2024.14101926

constructs an assessment framework for high school students' scientific modeling abilities in biology and attempts to develop assessment tools. Ultimately, the study uses the Rasch model to test these tools, demonstrating their validity and providing evaluation ideas for teachers and other researchers.

Keywords

Scientific Modeling Competence, Rasch Model, Assessment Study

Copyright © 2024 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

科学建模能力是高中教育阶段学生生物学科学思维的必备能力之一,其能力的培养与发展对全面落实学生生物学科核心素养有重要意义。根据文献研究,科学建模能力表现为当学习者面临问题情境时,通过调用头脑中的相关概念、建立概念关系网、构建心智模型,并以此为基础进行问题的解决和分析。由此可知,科学建模能力具有内隐性,如何将其外显进行评价,确立可操作化的评价标准和评价手段,是发展学生科学思维的关键。本研究旨在解决以下问题:如何科学有效地评价学生科学建模能力水平?如何检验测评工具是否真实反映出学生科学建模能力水平?

2. 科学建模能力测评框架

科学建模能力主要包括元建模知识和建模实践两部分,其中元建模知识是指关于建模过程本身的知识 and 理论,旨在帮助建模者更好地理解 and 应用建模方法。部分研究表明该部分能力水平与学生自身的科学建模能力水平相关。因此,有关于科学建模能力水平的评价研究主要是从元建模知识、建模实践或两者结合的角度出发。

2.1. 元建模知识测评框架

为了评估学生对于模型及其在科学教育中应用的理解,各学者开发了多种测评框架。Grosslight 等使用访谈法从模型种类、模型多样性、模型目的、设计和建构模型以及改变模型五个方面来调查学生对于模型本质的认识水平[1],并根据访谈结果将学生对于模型本质的认识水平划分为三个层次。Justi 和 Gilbert 在此基础上使用半结构式访谈对教师关于模型本质理解水平进行调查,增加了对模型本质理解水平的评价方法[2]。Treagust 等人通过对不同水平层次的学生进行随机抽样,并使用李克特五氏量表法调查其对科学模型的理解,最终根据调查结果将模型理解划分为五个方面:科学模型作为多重表征、模型是精确复制品、模型是解释工具、科学模型如何使用、科学模型本质在不断变化[3]。Kruger 等人认为模型是对事物阐明解释或作为一项研究工具验证想法、得出结论,并结合前人研究结果提出了适合于生物学科元建模知识测评框架(如图 1 所示)。

从模型的本体论和认识论出发,将模型理解划分为模型本质和模型多样性,反映模型本身描述和简化现象的方式。其中,模型多样性不单指模型的种类,更重要的是指模型能够反映客体事物的不同方面或以不同方式反映同一客体。从认知过程以及模型使用方式,将模型理解划分为模型目的、模型的检验和修改,反映模型本身可用来检验假设、预测未来事件和交流想法。Kruger 还进一步对理论框架各个组

成部分进行深化, 将其从低到高划分为三个水平层次。随后, 该理论框架被认为是科学有效的。

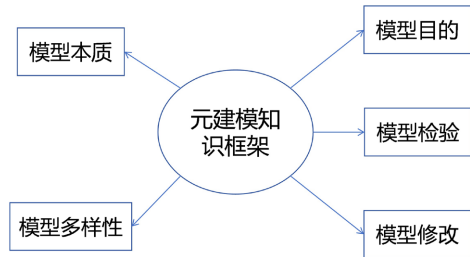


Figure 1. Kruger *et al.*'s theoretical framework for understanding models
图 1. Kruger 等人关于模型理解的理论框架

在我国《普通高中生物学课程标准(2017 年版 2020 年修订)》(以下简称《课标》)学业质量部分对学生使用模型和建模的能力提出了不同水平层次的要求, 如水平二要求学生能基于特定事实, 以文字、图示的形式, 说明相关概念内涵; 水平四要求学生能基于特定事实, 采用模型建模等方法, 以恰当的形式进行阐释和论述。因此根据《课标》要求并结合文献研究, 本研究将模型理解划分为模型本质、模型多样性、模型评鉴和模型应用四个维度, 并从低到高将每个维度划分为三个水平(如表 1 所示)。

Table 1. Assessment framework of metamodeling knowledge
表 1. 元建模知识测评框架

水平层次	评价维度			
	模型本质	模型多样性	模型评鉴	模型目的
水平一	模型是对原始客体的复制	模型只能表征一个原始客体	评鉴模型本身, 对模型自身错误进行修改	模型被用来描述原始客体
水平二	模型是对原始客体特征及结构的理想化表达	不同的模型能够表征同一原始客体	将模型与原始客体比较, 依据对于原始客体的变化进行修改	模型被用来描述和解释原始客体
水平三	模型是对原始客体特征及结构的形象表达或理论重构	不同的模型能够表征同一原始客体, 且关于原型事物的假设不同	用模型检验关于原始客体的假设, 基于检验结果修改模型	模型被用来描述、解释原始客体, 并预测其发展规律

不同维度的每个水平层次对应学生相应的行为表现。如在模型多样性维度下, 对其理解水平达到水平二的学生会认为同一个原始客体能够使用不同的模型进行表达。例如, 真核细胞的结构可以使用结构模式图表达, 也可以使用三维结构模型表达。因此, 评价学生的元建模能力可针对不同维度及水平层次设计相应的访谈问题或问卷量表等。

2.2. 建模实践测评框架

建模是科学知识构建和应用的系统活动, 而非字面意义的模型建立过程。Halloun 认为建模包括模型选择、模型建立、模型验证、模型分析和模型调度[4], 且中间三个环节是相互重叠、可同时进行。从模型中建构科学知识、形成心智模型, 再通过模型的建立呈现来表达心智模型。邱美虹认为建模过程具有序列性, 通过一系列过程让学生将日常的心智模式转化成科学模型, 并以 Halloun 建模过程为基础提出新的建模评价框架(如图 2 所示), 并根据 SOLO 分类评价理论和 Hempel 科学理论结构对建模过程进行水平层次划分, 将建模能力划分为六水平能力层次并构建了建模能力分析指标[5] (MAAI)。如处于经验反应层次的学生在面临问题情境时会聚焦于非理论的经验观察, 强调直觉经验, 无法说出与理论相关的因素,

进而无法做出合理的模型选择。其中，科学理论层次(Level 6)为建模能力最高层次，能够形成几乎完整的科学理论，并思考理论的价值，同时进行假设验证。张志康等人对 MAAI 进行实证研究后发现，几乎没有学生能够达到科学理论水平，大部分学生的能力水平集中于多重因素层次(Level 2)和关系层次(Level 3)之间。这为本研究建立建模实践能力水平划分提供了参考依据。

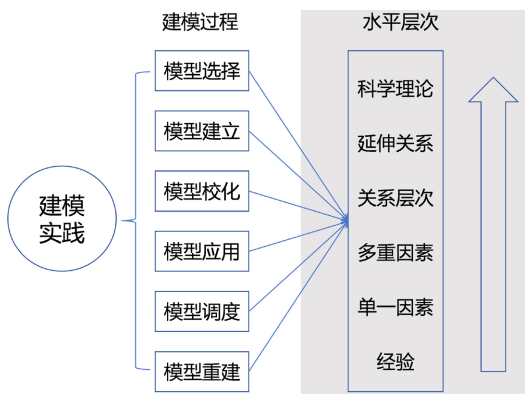


Figure 2. Modeling process and level division [5]
图 2. 建模过程及水平层次划分[5]

在我国《课标》中对于建模能力的发展虽未明确提出水平要求，但在模型建构、模型应用方面清晰指出了学生应当发展的能力。如要求学生能够建构模型如真核细胞结构模型、DNA 双螺旋结构模型、种群数量变化模型等，并要求学生能够结合模型描述生命规律及现象，并作出相应决策。因此根据《课标》要求并结合文献研究，本研究将建模实践划分为模型选择、模型建立、模型效化和模型应用四个维度，并从低到高划分为三个水平(如表 2 所示)。

Table 2. Assessment framework of modeling practice
表 2. 建模实践测评框架

水平层次	评价维度			
	模型选择	模型建立	模型效化	模型应用
水平一	无法反映出相关概念	无模型建立	仅凭经验判定，无考量因素	无应用
水平二	反映出与问题情境相关的概念	模型能够反映客体特征	利用各因素对模型进行评估	使用已效化模型中的单一结构解决相似情境中的问题
水平三	能够指出概念间关系的交互作用，进一步延伸和抽象	模型能够以抽象形式反映原始客体	利用多重因素间因果关系的交互作用对模型进行评估	使用已效化模型中的关系结构的交互作用解决相似情境中的问题

3. 测评工具开发与核验

3.1. 测评工具开发

根据文献研究结果，对于科学建模能力测评大多采用访谈法、问卷调查法以及纸笔测验法，少数研究者通过计算机程序进行测评研究。由于科学建模能力内隐性的特点，在使用问卷调查法以及纸笔测验时常配合访谈法使用，以便更好检测测评工具的有效性和更准确地反映被测者的能力真实水平。因此，本研究采用纸笔测验配合访谈法的方式对学生进行科学建模能力水平测评。

依据《课标》要求，试题设计要从生物学情境出发、贴近生物学事实。《课标》明确提出学生要能够

基于生物事实并运用模型与建模的方式阐释和揭示生命现象和规律。换言之, 试题情境能够贴近学生了解或者熟知的相关生物学现象, 便于学生提取关键信息并作出判断。

[例 1]某生物小组对某疾病发病机制进行研究时, 提出假设“该疾病是由于某种病毒入侵细胞从而导致发病”, 并制作了模型对该假设进行解释说明。你如何评鉴该模型?

本题测评元建模知识维度下学生对模型评鉴的理解水平。通过具体生物学情境, 学生以已有认知结构中的相关概念或理论为基础, 进而评价模型的科学性和有效性。比如该模型能否科学地解释该疾病发病机制? 该模型能否反映出该疾病所有致病因素? 通过学生对模型不同的评鉴方式, 进而判断其模型评鉴的能力水平。

应当注意的是, 针对建模实践维度进行试题编制时, 应避免先验知识的影响。建模实践测评一般从建模过程及建模结果两个角度进行, 过于熟悉的情境特别是直接搬用教材内容如“请构建动植物生命层次概念模型”, 会让学生直接从认知结构中调动已有模型或受到已有模型的干扰, 无法了解学生真实的水平能力。由此, 测评建模实践的能力水平应当从学生的最近发展区出发, 以已有的科学知识为基础, 构建新的生物学情境。

科学建模能力是多维度多水平的能力结构, 试题设计需全维度全水平覆盖。换言之, 针对封闭式项目, 每个维度至少设计 3 个项目, 且每个项目选项应当覆盖该能力的所有水平层次, 避免选项过高或者过低反映学生能力水平。如果是采用李克特量表进行调查, 则要确保项目既有正向描述也有反向描述。通过科学合理设计项目选项, 以便学生的选择进而判断其能力水平层次。

[例 2]电镜下观察到的细胞膜结构与教材呈现的细胞膜磷脂双分子层结构模式图之间有何联系? ()

- A. 精确地反映了细胞膜的真实结构
- B. 是对细胞膜真实结构的形象表达, 反映了细胞膜结构特征
- C. 是科学家基于对细胞膜的真实发现, 对其结构特征进行简化抽象所得到的

本题测评元建模知识维度下学生对模型本质的理解水平。依据选项 - 能力水平相对应的原则, 通过题目作答情况反映学生对于模型本质的理解水平, 比如选择 A 的学生认为模型是精确的复制品, 是对客体的准确反映, 进而判断该学生对于模型本质的理解应当处于水平一。

除上述以外, 试题编制还需尽可能避免学生猜题作答。试题选项之间的差异不明显、某个选项描述过长或者使用生僻词汇, 会导致学生出现猜测答题现象, 进而影响测评结果。

本研究试题共 16 个项目, 测试项目与评价维度的对应情况如表 3 所示。测评工具初次编制后, 经研究小组多次讨论和专家问询对测评工具进行修改删减, 确保其效度。

Table 3. Correspondence between test items and framework indicators

表 3. 测试题项与框架指标对应情况

评价维度		题项
元建模知识	模型本质	1、2、3
	模型多样性	4、5、6
	模型评鉴	7、8、9
	模型目的	10、11、12
建模实践		13、14、15、16

3.2. 测评工具核校

研究对象为某市普通高中高二年级选科生物的两个班 84 人, 发放测试卷 84 份, 回收 84 份, 回收率为 100%。为了更加科学有效地检验测评工具质量, 本研究利用 Winsteps3.7.2 软件基于 Rasch 测量模型

对测试数据进行分析，并根据分析结果来决定是否需要为题项进行进一步的修正。Rasch 模型是基于项目反映理论(IRT)发展出的验证数学模型，在心理学领域中常被用来检测测量工具的质量好坏。项目反映理论以概率函数的形式来描述项目作答反应的结果，且该结果只受被试能力水平与项目的计量学属性有关，并能够将被试的表现和项目难度定位在同一个潜在变量上，从而可以直接比较这两个参数。相比之下，经典测量理论是以被试的一般表现即观测分数对测量工具进行评价，测量工具的好坏是依靠所选被试的能力水平决定，反之被试能力水平的高低则是通过使用该测量工具得到的观测分数来评价。显而易见，经典测量理论具有参数依赖的局限性，而由 IRT 理论发展出来的 Rasch 能够解决该问题。利用 Rasch 模型对测评工具质量进行检验常通过拟合度、难度、区分度、信度、单维性分析、怀特图以及气泡图等方式进行。

3.2.1. 总体质量分析

经过测试结果分析(如表 4 所示)，测量工具试题信度 0.97 (>0.7)符合测量工具的信度要求。其中，Infit 和 Outfit 反映该测试数据与 Rasch 模型的拟合程度。在理想状态下，MNSQ 值为 1，ZSTD 值为 0，越趋近于理想值表明该工具质量越好。测评工具的总体拟合情况如表 4 所示，均趋近于理想值，且在可接受范围(-0.5~1.5)内。因此，该测量工具整体质量状况较好，可进行后续分析。根据分析结果，题项的分离度为 5.51 (>2)，表明测试题目分布较广，可以覆盖不同能力水平，而被试的分离度为 1.28，说明该测试项目对于被试的区分度不是很好。

Table 4. Overall quality analysis of assessment tools
表 4. 测评工具总体质量分析情况

	Infit		Outfit		Separation	Reliability
	MNSQ	ZSTD	MNSQ	ZSTD		
题项	0.97	-0.4	0.99	-0.2	5.51	0.97
被试	1.00	0	0.99	0	1.28	0.70

3.2.2. 单维性分析

单维性检验旨在确认测量工具是否只测量一个特定的心理特征或能力，而不是多个不同的特征。换言之，该测试题的作答结果只与被试某一种能力有关。一般认为该相关关系值在-0.4~0.4 之间，则表明测试项目符合单维性要求。通过单维性分析得到该测量工具的标准残差对比图(如图 3 所示)。

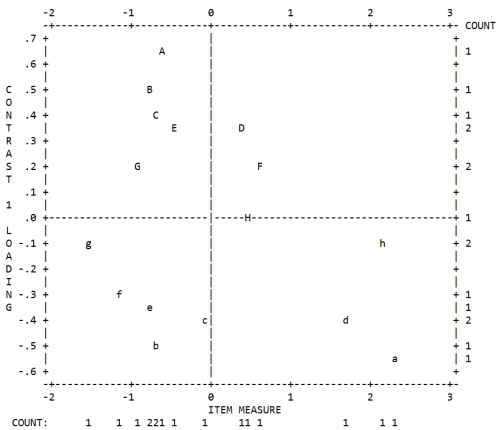


Figure 3. Standard residual comparison chart
图 3. 标准残差对比图

测评工具中的测试项目大部分处于-0.4~0.4 的范围, 有四个题项(2, 8, 11, 16)落在范围之外, 表明这四个题项可能涉及其他潜在因素即不满足单维性要求, 或题目本身措辞有误引起被试误解等情况。因此, 在对题项进行修改时有必要对学生进行访谈, 通过了解学生作答情况来判定题目本身是否会引起学生认知错误进而导致题项无法准确反映学生真实能力水平。

3.2.3. 怀特图

通过怀特图可粗略地判断题项的质量, 将被试能力水平与题项放置同一标尺, 进而判断试题难度是否匹配被试水平。如图 4 所示左侧反映被试能力水平的分布情况, 右侧反映测试项目的难度分布情况。由图可知, 题项难度分布均匀, 范围较广, 基本覆盖被试能力水平。被试的能力水平也呈正态分布, 中间多两边少。由此可见, 该测评工具难度适中, 可以很好地反映被试的能力水平。

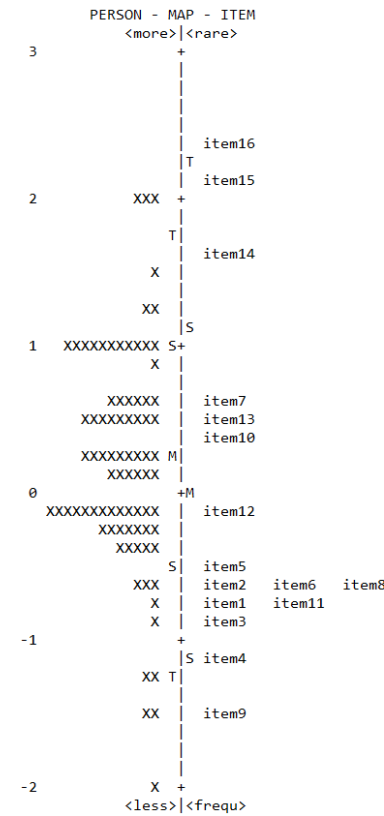


Figure 4. Wright map
图 4. 怀特图

3.2.4. 各题项质量分析

Rasch 模型是一个理想化的数学模型, 要求所收集到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[6]。如图 5 所示测量工具各题项与 Rasch 模型的拟合情况, 大部分题项拟合值均在 -0.5~1.5 之间, 但个别题项如 5、11 和 12 拟合值不好, 分别是 1.88、1.82 和 1.69。这意味着题项与被试能力不符, 可能存在能力水平高的学生在作答时未选择对应选项或者题目选项描述导致被试产生误解从而选择不符合自身水平的选项。此外, Rasch 标准误差表明测量工具的稳定性, 误差越小, 则测评工具在测量能力水平时的稳定性就越高, 即信度越高。从表格中可知题项的标准误差在 0.17~0.23 之间, 均处于可接受范围, 且题项的所有点 - 测量相关系数均在 0~1 之间, 表明该测量工具对能力的估计较为精确。

ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	ITEM
16	107	83	2.33	.23	.70	-1.8	.66	-1.6	.50	.32	75.9	73.8	item16
15	110	83	2.17	.22	.59	-2.8	.57	-2.4	.61	.33	88.0	69.8	item15
14	121	83	1.68	.20	.60	-3.1	.67	-2.2	.50	.36	72.3	61.4	item14
7	152	83	.63	.17	1.19	1.4	1.15	1.1	.52	.41	44.6	53.7	item7
13	158	83	.45	.17	.64	-3.2	.65	-3.1	.56	.42	67.5	54.3	item13
10	160	83	.39	.17	.79	-1.7	.82	-1.4	.23	.42	66.3	54.4	item10
12	176	83	-.08	.17	1.69	4.5	1.73	4.7	.32	.42	32.5	53.6	item12
5	189	83	-.48	.18	1.88	5.4	1.83	5.0	.27	.41	32.5	54.4	item5
2	194	83	-.65	.18	.85	-1.2	.89	-.7	.19	.41	53.0	55.0	item2
6	195	83	-.68	.18	.60	-3.5	.66	-2.7	.39	.41	68.7	54.9	item6
8	195	83	-.68	.18	.84	-1.2	.81	-1.4	.54	.41	66.3	54.9	item8
1	197	83	-.75	.18	.74	-2.1	.75	-1.9	.50	.40	60.2	54.7	item1
11	198	83	-.78	.18	1.82	4.9	1.86	4.9	.35	.40	36.1	56.1	item11
3	202	83	-.92	.19	.71	-2.3	.71	-2.1	.60	.40	65.1	56.4	item3
4	208	83	-1.14	.20	.94	-.4	1.09	.6	.19	.39	54.2	59.8	item4
9	217	83	-1.51	.21	1.00	.0	1.03	.2	.27	.36	66.3	66.9	item9
MEAN	173.7	83.0	.00	.19	.97	-.4	.99	-.2			59.3	58.4	
S.D.	34.2	.0	1.14	.02	.42	2.9	.42	2.7			15.5	6.1	

Figure 5. Fit indicators for the 16 items
图 5. 16 个项目拟合分析

4. 小结

由于科学建模能力的内隐性，本研究旨在为评价科学建模能力提供具有可操作性的思路，初步构建了适合我国高中生物学的科学建模能力评价框架，但其各维度及水平层次划分的科学性与准确性还有待后续研究的验证。此外，由于传统评价测评工具质量好坏的方式存在一定局限性，本研究为此提供新的验证方法——基于项目反映理论开发的 Rasch 模型——适合于任何测评工具质量的检验，能够帮助教师在日常教学工作中优化评价方式，改进评价工具质量。

参考文献

[1] Grosslight, L., Unger, C., Jay, E. and Smith, C.L. (1991) Understanding Models and Their Use in Science: Conceptions of Middle and High School Students and Experts. *Journal of Research in Science Teaching*, **28**, 799-822. <https://doi.org/10.1002/tea.3660280907>

[2] Justi, R. and Gilbert, J. (2003) Teachers' Views on the Nature of Models. *International Journal of Science Education*, **25**, 1369-1386. <https://doi.org/10.1080/0950069032000070324>

[3] Treagust, D.F., Chittleborough, G. and Mamiala, T.L. (2002) Students' Understanding of the Role of Scientific Models in Learning Science. *International Journal of Science Education*, **24**, 357-368. <https://doi.org/10.1080/09500690110066485>

[4] Halloun, I. (1996) Schematic Modeling for Meaningful Learning of Physics. *Journal of Research in Science Teaching*, **33**, 1019-1041. [https://doi.org/10.1002/\(sici\)1098-2736\(199611\)33:9<1019::aid-tea4>3.0.co;2-i](https://doi.org/10.1002/(sici)1098-2736(199611)33:9<1019::aid-tea4>3.0.co;2-i)

[5] 张志康, 邱美虹. 建模能力分析指标的发展与应用——以电化学为例[J]. 科学教育, 2009, 17(4): 319-342.

[6] 晏子. 心理科学领域内的客观测量——Rasch 模型之特点及发展趋势[J]. 心理科学进展, 2010, 18(8): 1298-1305.