

# 基于 Rasch 模型的化学和物理学跨学科能力测评研究

研究生姓名：石超

导师姓名：许燕红

学科：学科教学（化学）

研究方向：化学教育

年级：2022 级

## 中文摘要

跨学科能力，作为一种综合性极强的关键能力，随着新课程改革的进一步深化，跨学科能力的培养和评价受到越来越多的关注。然而，当前针对高中生跨学科能力的测评工具及其相关研究较少，因此，本研究旨在开发相关测评工具并探究高中生跨学科的能力水平分布为一线教师提供参考。

在分析 Rasch 模型、跨学科能力等相关资料的基础上，明确研究的目的、方法和思路，并对跨学科能力概念进行界定，确定了化学与物理跨学科能力测评的要素：信息的提取与转译能力、知识的整合与联系能力、概念的融合与量化能力、问题的解决能力。针对这些要素，将每种能力划分为两个水平，并详细描述了各水平的具体表现，构建了跨学科能力各要素的测评框架。根据构建的测评框架，针对初、高中物理、化学学科知识点编制了测评的初试试卷，并进行试测。再以 Rasch 模型为指导，运用 Winsteps3.72.3 软件对数据进行深入分析，获得了整体质量统计、项目拟合和误差统计、标准残差图、怀特图等关键信息，从而全面检验了试题的信度和效度。并根据试测结果对测评工具进行了优化，再运用该工具对桂林市三个中学的高三学生进行了跨学科能力测评，分析得出高三学生化学和物理跨学科能力的现状。

研究结果显示：所编制的测评工具质量良好，被试与项目的信度和分离度均表现出色，拟合度参数接近理想值，误差较小。桂林市中学的高三学生在信息的提取与转译能力方面表现较强，主要处于水平二；而在知识的整合与迁移能力、概念的融合与量化能力和问题的解决能力方面则主要处于水平一，相对较为一般。基于上述研究结果，提出以下教学建议：注重审题教学，明确题目关键信息；注重知识迁移，构建完整的知识体系；注重定量研究，加强计算能力；注重设计问题情境，培养问题解决能力。

尽管本研究取得了一定成果，但样本量容量偏小，对高三学生跨学科能力水平分布的探究尚不够广泛。后续研究可进一步扩大样本量和覆盖面，进一步修订能力表现标准，完善测评工具。

**关键词：**Rasch 模型；跨学科能力；能力测评

# **Research on Interdisciplinary Competence Assessment of Chemistry and Physics Based on Rasch Model**

Graduate student: Shi chao

Supervisor: Xu yanhong

Major: Subject Teaching(chemistry)

Research direction: Chemical Education

Grade: 2022

## **Abstract**

Interdisciplinary ability, as a highly comprehensive key ability, has received increasing attention in the cultivation and evaluation of interdisciplinary ability with the further deepening of the new curriculum reform. However, there are currently few assessment tools and related research on the interdisciplinary ability of high school students. Therefore, this study aims to develop relevant assessment tools and explore the distribution of interdisciplinary ability levels among high school students to provide reference for frontline teachers.

Through literature research, this article defines interdisciplinary competence as the ability of students to comprehensively and holistically grasp complex scientific challenges by applying knowledge and thinking from multiple disciplines such as chemistry and physics. At the same time, interdisciplinary competency elements are divided into: ability to extract and translate information, ability to integrate and connect knowledge, ability to integrate and quantify concepts, and ability to solve problems. In response to these elements, this study divides each ability into two levels and provides a detailed description of the specific performance of each level, constructing an evaluation framework for each element of interdisciplinary ability. On this basis, this study developed assessment papers for the knowledge points of physics and chemistry in middle and high schools, and conducted interdisciplinary ability assessments on senior high school students from three high schools in Guilin. Guided by the Rasch model, Winsteps 3.72.3 software was used to conduct in-depth analysis of the data, obtaining key information such as overall quality statistics, item fitting and error statistics, standard residual plots, White plots, etc., thus comprehensively testing the reliability and validity of the test questions. Based on the test results, the evaluation tool was optimized and used to deeply explore the current situation of interdisciplinary abilities among high school seniors.

The research results show that the quality of the evaluation tool developed is good, and

the reliability and separation between the subjects and the project are excellent. The fitting parameter is close to the ideal value, and the error is small. High school seniors in Guilin City have strong abilities in information extraction and translation, mainly at level two; In terms of knowledge integration and transfer ability, concept integration and quantification ability, and problem-solving ability, it is mainly at level one and relatively average. Based on the above research results, the following teaching suggestions are proposed: pay attention to teaching by reviewing questions and clarify the key information of the questions; Emphasize knowledge transfer and build a complete knowledge system; Emphasize quantitative research and strengthen computational capabilities; Emphasize the design of problem scenarios and cultivate problem-solving abilities.

**Key Words:** Rasch model; Interdisciplinary competence; Ability assessment

# 目录

中文摘要.....	I
Abstract.....	II
目录.....	IV
1 绪论.....	1
1.1 研究背景.....	1
1.1.1 基于高中化学课程标准的要求.....	1
1.1.2 基于化学与物理的学科特征.....	1
1.1.3 跨学科能力的现实意义.....	2
1.2 研究现状.....	2
1.2.1 Rasch 模型研究.....	2
1.2.1 跨学科能力研究.....	4
1.2.3 化学与物理跨学科研究的现状.....	9
1.2.4 相关研究评述.....	10
1.3 研究目的和意义.....	11
1.3.1 研究目的.....	11
1.3.2 研究意义.....	11
1.4 研究方法.....	12
1.5 研究思路.....	12
2 概念界定与理论基础.....	14
2.1 跨学科能力的界定.....	14
2.2 理论基础.....	15
2.2.1 学习迁移理论.....	15
2.2.2 多元智力理论.....	15
2.2.3 建构主义学习理论.....	16
2.2.4 项目反应理论.....	16
3 化学与物理跨学科能力测评框架的建构.....	18
3.1 化学与物理跨学科能力的组成要素.....	18
3.1.1 文献中的跨学科能力模型.....	18
3.1.2 化学与物理跨学科能力组成要素.....	19
3.2 化学与物理跨学科能力模型的建构.....	22
3.2.1 信息的提取与转译的水平建构.....	23
3.2.2 知识的整合与联系的水平建构.....	23
3.2.3 概念的融合与量化的水平建构.....	23
3.3.4 问题的解决的水平建构.....	24
4 测评工具的设计和优化.....	25
4.1 测评工具的开发流程.....	25
4.2 测评工具的开发与设计.....	25
4.3 测评工具的试测和优化.....	31
4.3.1 样本选取和数据收集.....	31
4.3.2 各子能力数据收集.....	31
4.3.1 测试工具的优化.....	48

5 化学与物理跨学科能力测评分析 .....	51
5.1 整体质量分析 .....	51
5.2 信息的提取与转译能力数据分析 .....	54
5.3 知识的整合与联系能力数据分析 .....	58
5.4 概念的融合与量化能力数据分析 .....	62
5.5 问题的解决能力数据分析 .....	66
6 研究结论与建议 .....	71
6.1 研究结论 .....	71
6.2 教学建议 .....	71
6.3 研究反思与展望 .....	73
参考文献 .....	74
附录 .....	78
附录一高三跨学科能力测试卷（试测） .....	78
附录二高三跨学科能力测试卷（最终卷）及答案 .....	83
致谢	
论文独创性声明	
论文使用授权声明	

# 1 绪论

## 1.1 研究背景

### 1.1.1 基于高中化学课程标准的要求

在 2017 年公布且于 2020 年更新的《普通高中化学课程标准》中，强调了化学作为一门综合性较强的学科，它既与传统的学科如物理、生物和数学有着密切的联系，也与工业界的多个方向比如材料科学、环境科技、信息技术等领域紧密相关。其中还特别指出，在教学过程中应加大对化学与其他学科融合的重视，鼓励在化学教育中采纳跨学科的教学模式。教师在组织课程内容时，应注重如何把化学知识与其他领域的知识相结合，从而帮助学生建立起更加立体和全面的知识架构。这样的教学方式能够极大地拓展学生的视野，启迪他们的创新能力，并指导他们学会综合应用各学科知识来解决实际问题。这种跨学科的教学策略不仅与新课程标准的推荐理念相吻合，也有助于学生更扎实地掌握科学知识，促进他们的综合素质和核心能力的发展。简言之，通过采用跨学科的教学方法，可以有效地促进知识的整合，激发学生的创新意识，同时，还能够帮助他们构建坚实而全面的知识体系，为面对复杂问题提供更为广阔的视角和多元的解决方案<sup>[1]</sup>。

### 1.1.2 基于化学与物理的学科特征

作为自然科学的一部分，物理、化学、生物之间存在着许多相通之处。特别是在物理和化学之间，这种联系更为紧密，科学史上，许多科学家既是化学家又是物理学家。因此，不应该将物理和化学看作是两个完全不同的学科，而是应该将它们视为自然科学的两个方面。在物理课程标准的核心素养框架内，“科学思维”具体体现在构建物理模型、逻辑推理、实验验证以及提出科学问题等多个重要环节。这种素养要求学生具备深刻理解和把握客观世界事物属性、自然规律及其相互间关系的能力。此外，科学思维还涵盖了对自然现象进行深入探究以及解决实际物理问题的能力，它要求学生能够熟练运用科学方法进行逻辑推理、分析和论证，同时保持对不同观点和结论的开放性和批判性，勇于提出质疑并不断完善修正<sup>[2]</sup>。这与化学中的证据推理、模型认知和创新意识的核心思想有许多共通之处。在这些科目中，学生不仅需要熟练掌握科学探究和科学态度等方面的知识，还要掌握不同学科的知识和研究方法，并能够将他们应用到其他学科中，那么他们的学习迁移能力就会得到提高，进而发展学生的综合能力，而综合能力的发展，需要融合不同的学科、进行跨学科教学<sup>[3]</sup>。虽然在两科中许多知识点都相互联系，但是许多知识的表征形式和学习方式都是有所区别的。比如，在物理中，通常会使用数学方法和实验来推断和验证某些理论，而在化学中，更加强调通过实验和实践获得更多的感性认识，然后结合自身的知识和经验，进行科学抽象分析，上升到理性认识。所以，在教学过程中，只有把物理和化学的学习方式融合到教学过程中，让学生接触不同的学习方

式，利用不同的分析方法，通过系统地学习和理解，才能完全掌握知识的全貌，掌握更多的学习方式，并建立起完整的知识体系，避免在某些领域产生偏见。通过跨学科的学习，可以建立起综合的科学观念，发展学生的综合能力，从而提高他们对自然科学的理解和应用水平。因此，应该鼓励学生在学习分科知识的同时，积极参与跨学科的学习和探索，以达到更好地学习效果<sup>[4]</sup>。

### 1.1.3 跨学科能力的现实意义

在近年来的全球化发展和经济迅速增长的背景下，我国已跃居为全球第二大经济体。然而，伴随而来的社会与环境问题也日趋严峻，例如气候变化、环境污染、能源使用以及健康医疗等问题。这些问题的解决不仅需要自然科学知识，还需要数学、经济学、社会科学等多学科的知识技能。因此，跨学科教育已成为全球教育领域的发展趋势<sup>[5]</sup>。

在我国，长期以来的分科教学模式使学生在单一学科的知识点上表现出色，但在跨学科能力方面却有所欠缺。这种教学模式导致学科间的联系被割裂，影响了学生的综合思考和问题解决能力。因此，培养跨学科能力已成为国内外教育改革发展的方向之一。实践证明，跨学科教育不仅能帮助学生掌握综合知识，更有助于提高他们在面对复杂问题时的思考和解决能力。此外，跨学科教育还能提高学生的综合素质和创新能力，为其未来的职业生涯发展奠定坚实基础<sup>[6]</sup>。

## 1.2 研究现状

### 1.2.1 Rasch 模型研究

#### (1) 国外研究现状

在中国知网上以“Rasch 模型”为主题进行外文文献检索，研究的主要领域为用于对学生某些能力的测评，测评工具一般用自编试卷。

2007 年，Chandrasegaran A L 设计了一种二段式问卷，并采用了 Rasch 模型作为工具，旨在评估学生对于化学反应的宏微结合表征能力。经过深入研究，结果显示，学生能够利用语言描述化学变化的宏观现象，并且从微观角度解释这些宏观现象的成因。同时，学生也能够使用合适的化学式和化学反应方程式等手段对化学反应过程进行准确地表征<sup>[7]</sup>。

2016 年，William L. Romine 等研究人员研发出一款新的测评量表，旨在全面评估本科生酸碱概念的学习情况。为了确保量表的质量水平达标，他们借助 Rasch 模型对测评量表进行了评价。经过数据处理，发现大部分本科生并未完全掌握 pH 值标准的相关知识。这一结果表明在酸碱概念的教学中，需要进一步加强 pH 值标准相关知识的传授和巩固<sup>[8]</sup>。

2020 年，Erguvan 对学生的第一个期中考试论文进行了评分，并与研究人员分享了分数。此外，从每个班随机抽取 2 名学生，所有教师共标记 6 篇论文作为锚定目的。采

用多面 Rasch 模型 (MFRM) 进行数据分析。结果显示, 尽管评分者在所有考生和任务的评分中使用的规则一致, 但他们的宽容程度有所不同<sup>[9]</sup>。

## (2) 国内研究现状

Rasch 模型在国内的应用主要涵盖教育学、心理学、医疗和人事测评等多个领域。为了更深入地探索 Rasch 模型在化学教育中的实际应用情况, 以“Rasch 模型”和“化学”为主题在中国知网检索, 文献总数为 227 篇, 如图 1-1 所示, 根据研究内容分析, 主要的研究方向可以分为能力测评和试卷分析。

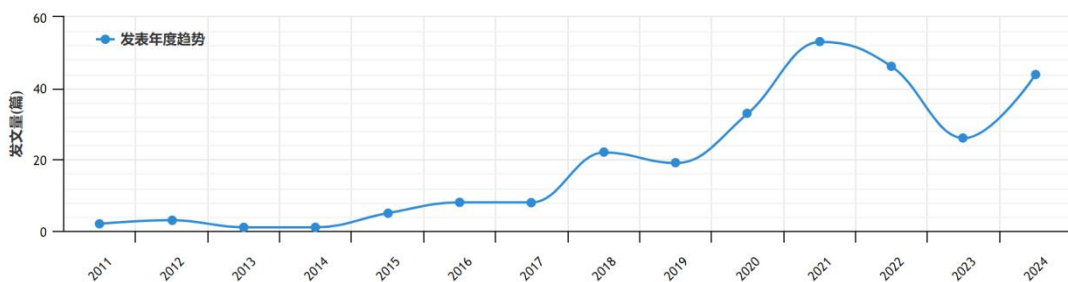


图 1-1 “Rasch 模型”和“化学”结合的研究发文量趋势图

关于能力测评方面的研究, 2012 年, 王祖浩对高一、高二及高三学生的“化学实验认知能力”进行了详尽的测评, 旨在了解这一能力在不同年级的分布情况以及学生的“模型认知”现状。通过深入分析, 揭示了“化学实验认知能力”的发展特征, 并进一步探讨了不同年级以及学校之间学生在化学实验能力方面存在的差异<sup>[10]</sup>。

2023 年, 梁晨在研究了相关文献和中学化学课程后, 明确了化学元素观的定义, 并将其划分为三个维度: 元素基本认知、物质视角的元素理解和化学反应视角的元素理解。为了评估高中生的化学元素观水平, 他基于 Chemquery 评价系统将其划分为四个水平, 并开发了相应的测评工具。通过测试, 他发现高中生的化学元素观主要集中在水平 2 和 3, 且大部分学生在基本认知维度上表现较弱。因此, 他建议教师在教学中以化学元素观为指导, 整合具体知识, 强化元素化合物知识的学习, 增进对元素周期律的理解, 并实施“教-学-评”一体化策略, 以提高学生的化学元素观水平<sup>[11]</sup>。

2023 年, 韩泽宇对初中毕业生的科学推理能力进行了全面的评估。经过严谨地分析, 他发现这些学生的整体能力水平略高于所使用测试题的难度标准。然而, 也注意到他们尚不能将科学推理技巧与数学运算进行有效地结合, 并且在控制实验变量、相关关系推理以及假设演绎推理等方面的表现还有待提高。因此, 建议在未来的教学中, 加强对这些方面的训练和指导, 以提升初中毕业生的科学推理能力<sup>[12]</sup>。

关于试卷分析方面的研究, 2016 年, 王桂桃采用 Rasch 模型, 对初三学生的中考化学试题进行了详尽的分析。经过深入研究, 他发现这套试卷的信度表现优异, 具备出色的区分度, 与 Rasch 模型有着良好的拟合效果。然而, 他也指出这套试题的难度略显简单, 较之受试者的能力水平略显不足, 尤其缺乏更具挑战性的题目。基于这些发现, 王

桂桃提出了一些可能存在的问题以及相应的解决策略，对于进一步优化试题分析方法具有重要的指导意义<sup>[13]</sup>。

2021年，徐佳敏采用 Rasch 模型，对上海市某区全体高三学生的一模试题进行了深入分析。研究结果显示，该测试卷的信度和区分度表现良好，试题的平均难度略低于学生的平均能力水平。在试题的整体单维性方面，大部分试题符合要求，但部分试题的拟合性有待提高。为了构建更有效的测评体系，徐佳敏从试题设计、评分细则以及教学方法三个方面给出了具体建议<sup>[14]</sup>。

经过深入研究 Rasch 模型的相关文献，我们发现，基于 Rasch 模型的能力测评研究在教育界一直备受关注，并已形成了相当成熟的研究体系，为本研究的能力测评提供了坚实的理论基础和实践指导。

### 1.2.2 跨学科能力研究

#### (1) 国外研究现状

在中国知网上以“跨学科能力”为关键词进行外文文献检索。在国外跨学科这个概念已经引起了教育研究人员的关注，主要可以分为跨学科构成要素，跨学科能力的培养，跨学科能力的测评和跨学科教学研究为主。

关于跨学科能力构成要素方面的研究，Shen 和他的同事们在 2014 年提出了一个旨在帮助学生将不同学科知识高效融合的模式。该模型核心由三个阶段组成：互译、适应和创新。在互译阶段，学生需要利用不同领域的语言和表述方法来说明同一概念，这样做有助于促进学科之间的沟通与相互理解。第二阶段，适应，要求学生能够将一个领域的理论或概念应用到其他学科，从而达到跨学科的深入结合。最后的创新阶段，需要学生将单一学科的模式和理论推广到跨学科整体中，以实现知识的创新运用。学生在这一过程中将专注点从单一学科的理解扩展到多学科知识的整合，进而转向对知识结构的调整、应用，并最终培养出综合性思维。通过这样的训练，学生能够在复杂问题解决和创新方面，展现出卓越的跨学科能力<sup>[15]</sup>。

关于跨学科能力培养方面的研究，芬兰在教育领域的创新举措得以在 2014 年的《国家基础教育核心课程》中体现，当中重点推崇了将学习内容与现实生活现象紧密联系的跨学科学习方法。对学生而言，这种教学方式鼓励更主动的参与度，并使学习过程更加贴合生活实际。芬兰也对之前的跨学科主题进行了升级，进化成包含了七种不同面向的跨学科素养，涵盖了诸如批判性思维、文化认知、人际互动、个人情感表达、时间和自我管理、包容多元观点的阅读、运用信息沟通技术、职业与企业家技能、以及培养社会参与及可持续发展意识等方面。在具体实施上，芬兰的教育体系通过将这些跨学科素养切割为更小的学习单元，将它们嵌入传统学科目标之中，引导学生在多元环境下有效灵活地应用所学知识，这不仅体现了跨学科能力培养的方针，也贯穿了整个课程规划的过程<sup>[16]</sup>。

关于跨学科能力测评研究，2016年，Mihwa Park 利用“能量”这一跨学科概念，精心设计了一种测评工具，并成功地对365名大学生进行了大规模的测试。经过严谨的数据分析，发现该测量工具具有较高的信效度，能够准确反映受试者在该领域的实际水平。此外，对各个项目的难度进行了深入剖析后发现学生们在理解波、电、磁、原子结构以及与现代物理学领域相关的能量概念方面存在一定的困难<sup>[17]</sup>。

2018年，HyeSunYou 研发出了一款测评工具，该工具以“碳循环”为基础，横跨物理、化学、生物和地理等多个学科，用于评估高中和本科生的科学知识水平。如图1-2所示，测评形式主要采用传统的纸笔测试，同时融入了跨学科和独立学科的教学项目。这款工具不仅能有效地衡量学生在碳循环相关问题上的解决能力，还能深入挖掘学生在跨学科理解上的短板，为他们提供更有针对性的指导和建议<sup>[18]</sup>。

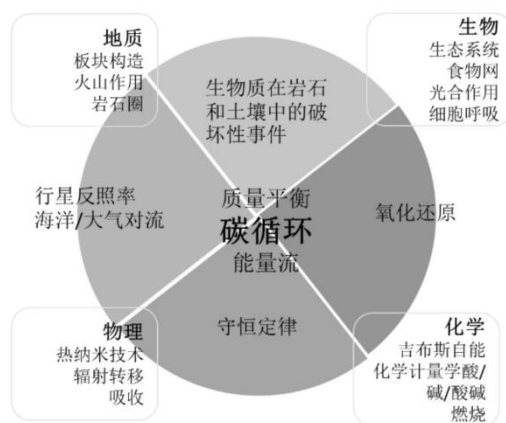


图 1-2 碳循环的跨学科关系

关于跨学科教学研究，最初由美国学者舒梅克所提出。此项研究致力于突破学科间的界限，将课程之多元面向有机整合，形成具有深度关联的知识体系。此种教学模式之目的，在于使学生能够于更宽广之学术领域中汲取知识，进而深化对现实世界的认知与理解。此外，跨学科教学亦重视教与学之紧密结合，以促进学生的全面发展<sup>[19]</sup>。

全球教育领域正经历跨学科教学方法的革新潮流，众多国家正在深入探索并实践这一先进的教育理念。美国作为早期探索者，在进化不断的教育系统中形成了一系列独特的教学模式，其中最为突出的是跨学科、多学科、以及超学科这三种课程设计。多学科课程充分考虑了不同学科的结合，它基于每个学科独立的教学内容，并试图在保持学科原有界限的前提下实现一定程度的整合。例如，它可能将一个学科的不同分支聚合在一起，形成较为完整的学习体系。在此基础上，融合课程进一步将各项技能、知识和态度纳入单一学科的教学之中，旨在丰富学科内容，使之更加符合实际应用。进一步地，跨学科课程则突破了传统学科边界，更侧重于探索和应用不同学科间的共性知识点。此模式下，教育者被赋予了整合和运用这些横跨各学科的知识任务，以此提高学生的综合思维和分析技能。尽管其目的是打破学科间的隔阂，但现实中的跨学科相通点有限，这

种整合往往还是部分性的，并且教学方式依旧以教师为主导。最终，超学科课程代表了对学生中心教育模式的极致追求，它通过基于项目及问题导向的方法，来鼓励学生主动探究、解决问题，实现充分的跨学科整合。在这一模式中，学生的个人兴趣、课程标准、地方资源被有机地结合在一起，促成独立或合作完成的项目产物或探究成果。此过程不仅涵盖了交叉学科的知识，还强调了生活技能的培养，尊重学生的独立思考和创新能力<sup>[20]</sup>。这三种教学模式各自的特点和比较可以参见具体的对照表格 1-1。

表 1-1 三种课程模式的比较

课程模式	课程中心	教学目标	融合程度
多学科课程	某一学科	单一学科知识	无跨学科界限
跨学科课程	学科之间的共有知识	跨学科概念和技能	部分跨学科
超学科课程	项目、活动、主题	生活实践技能	完全跨越学科界限

德国在 20 世纪 90 年代中期即开始提倡中小学阶段的跨学科教学。2004 年的新教学大纲对“跨学科教学”给出了明确定义：以某一学科为核心，选定特定主题，并整合其他学科知识进行深入探讨。这种教学方式着重于从多元视角审视问题，且所选主题与现实生活紧密相连。其核心目标在于培养学生的跨学科思维，使他们能够从不同学科背景出发分析问题。需要强调的是，虽然各学科知识在此模式下相互交织，但它们仍保持各自的独立体系和学科界限<sup>[21]</sup>。

## (2) 国内研究现状

在中国知网上以“跨学科能力”为关键词进行文献检索，总共只有 120 篇，如图 1-3 所示，根据研究内容可分为跨学能力构成要素，跨学科能力培养，跨学科试题和跨学科能力测评。

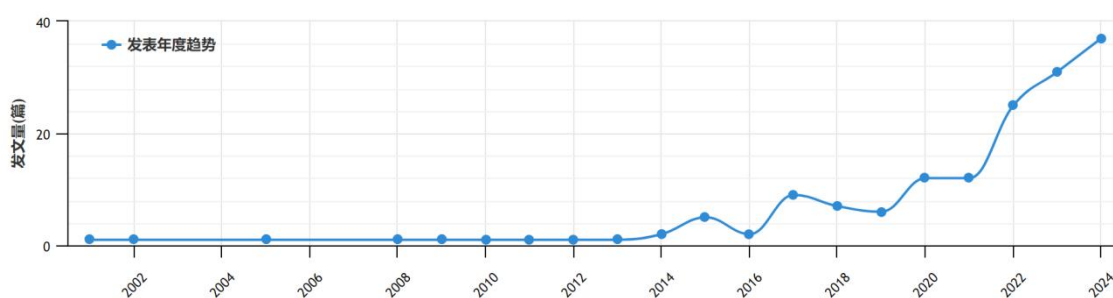


图 1-3 “跨学科能力”研究的发文量趋势图

关于跨学能力构成要素方面的研究，在 2017 年发表的《核心素养视角下的跨学科能力测评研究》一文中，王光明及其团队深入探讨了跨学科能力的核心构成。他们认为，这种能力主要包括三个方面：一是获取并理解知识的能力，二是将学到的知识应用到实际中去的能力，三是创造性地将知识应用于新场景的能力。通过这样的分类，研究旨在为跨学科学习提供一种评估框架，帮助理解和评价个体在跨学科环境中的表现<sup>[22]</sup>。

在 2020 年，上海市教育考试院推出了一份名为《初中相关课程终结性评价指南》的文件，强调了初中阶段学生必须在地理与生命科学等跨学科领域内展现出一定的能力。按照该指南，学生应拥有跨学科问题解决的思维模式，这包括能够主动投身于解答实际问题的过程之中。为了详细阐释这种能力，指南将其分为三个关键方面：对信息的获取及其处理、问题的分析及提出质疑以及对结论的解释及创新。这个分解过程是为了保证学生在其学习生涯的早期就能够全方位地成长，从而为他们未来的教育和生活提供坚实的支持。<sup>[23]</sup>。

关于跨学科能力培养实践方面的研究，2023 年，高凌蕊开展“设计和制作简易制氧机”这一跨学科项目的教学工作，精心设计了完整的教学流程。还对跨学科教学的路径和实践提出了宝贵的建议<sup>[24]</sup>。

2023 年，杜博老师以“原电池”为教学案例，通过设计一系列紧密相关的学习任务，结合先进的手持技术工具，深入探讨了如何实现电池的稳定运行以及提升电池效率的内在机制。在此过程中，他巧妙地运用了物理和化学领域已有的知识框架和理论模型，实施了跨学科的概念教学。这种教学方法不仅锻炼了学生从宏观和微观两个层面搜集和整理证据的能力，更提升了他们分析和解读证据的技巧，从而全面提升了学生的“证据推理与模型认知”学科核心素养<sup>[25]</sup>。

2024 年，黄臻等学者以“我是火箭推进剂设计师”为主题的跨学科实践教学设计，旨在培养学生化学观念、科学思维、科学探究与实践等关键能力，提升学生化学素养和责任感。教学设计分为三个环节，共 6 课时，学生通过学习理解火箭推进剂设计的基本原理，掌握跨学科知识解决实际问题的能力。教学过程注重培养学生的综合运用跨学科知识解决实际问题的能力，提升学生的实验探究能力和科学素养。学生通过项目分析与拆解活动，探讨了火箭升空、火箭推进剂选择、氢能源认识等问题，理解了化学原理，并运用知识分析燃料性质和应用。教师采用启发式、探究式和建构式的教学方式，引导学生开展高阶思维活动，帮助学生形成系统认识<sup>[26]</sup>。

2024 年，吉青青“一次性筷子的利与弊”为例，探讨了社会性科学议题教学。旨在培养学生解决问题、严谨逻辑思维、团队合作及知识建模能力。教学流程包括情境创设、任务指导、活动探究、成果展示和评价等环节。学生通过实验验证了硫磺和二氧化硫含量，并分析了其对人体的危害。实验报告单展示了学生的实验结果，提出了购买筷子时应选择正规厂家生产的建议。教师从实验目的、内容、用材、过程、结果及讨论等方面对学生进行了评价。教学实现了跨学科教学，提升了学生的动手能力、辩证思维和语言组织表达能力。但也发现活动开展存在不足，如可开展更多活动探讨一次性筷子对伦理道德的影响，以及创新活动形式和评价机制<sup>[27]</sup>。

关于跨学科能力试题方面的研究，跨学科的试题主要出现在部分省市的自主命题的高考题、自主招生题和竞赛题等方面。2013 年，王俊玲针对高考中的化学及文科与理科

综合题进行了深入剖析，并指出综合题已成为高考的必然趋势。她认为，这种题型能够有效地提高学生的综合运用知识解决问题的能力<sup>[28]</sup>。

2022年，牛宝妮对历年高考化学与其他学科的交叉融合进行了深入分析，并针对此现象提出了一些教学建议。她强调，高三化学教师应抓住这些知识点的内在联系，着重培养学生的综合素质和创新思维能力<sup>[29]</sup>。

关于跨学能力测评方面的研究，在2019年，宋歌开发出一款评估工具，专门用来衡量中学生在多门学科上的综合素质，尤其是针对初中九年级的学生群体。该项测评涉及了385名学生，这些学生分布在两个实验型班级、四个重点班级以及两个普通班级中。本项工具旨在通过四个关键维度——多元分析、整合知识、科学辩证以及实验研究——来评测学生们的跨学科能力。通过对知识整合理论的深入研究，宋歌提出了一个包含三个层级的知识整合模型，每个层级由学生行为的详尽描述组成，具体的对照表格1-2。为了确保评估结果的准确性，各个项目都与这些层级紧密对应。评估过程采取分级打分制，根据学生回答的具体内容进行评分编码。其后，通过使用Rasch模型对评估工具进行校验，该工具被证明在可靠性和效度上都表现良好。作者还运用了SPSS22.0这一统计软件，对知识整合维度的数据进行了深入的统计分析。研究发现，绝大多数学生(96.4%)的表现在第二个层级或更低层级。在性别方面，虽然男生的整体表现略好于女生，但两性之间并没有显著的差异。而在学业成绩方面，成绩优秀的学生比普通和成绩较低的学生展现出更高级别的知识整合能力。具体来说，优秀学生中有79.4%达到了第二层级，这一比率比整体样本要高。对于平均水平的学生，他们的水平分布与整体相当。然而，学业有困难的学生大多数落在第二层级或更低，其中79%的学生处于最低层次或以下。研究还发现，不同成绩水平的学生群体之间在知识整合水平上有显著的差异<sup>[30]</sup>。

表 1-2 知识整合水平的划分

知识整合	水平一	学生使用多学科知识解决问题，但这些知识间是彼此孤立的。
	水平二	学生建立起学科事实、概念、理论或方法之间的网络化关系
	水平三	学生将“整合”的知识迁移至另一全新情境，实现问题解决

2021年，宋琳以氧循环主题，对于跨学科理解的评估研究。他精心构建了氧循环的框架体系，并开发出一套科学的测评工具，用于对初中、高中一年级和二年级的学生进行测评。经过细致分析，他发现初中和高中一年级学生在氧循环主题的跨学科理解上并无显著差异。然而，高中二年级的学生在此领域的理解明显优于初中和高中一年级的学生，显示出显著的差异性<sup>[31]</sup>。

2022年，陈依捷深入研究了化学与生命科学的融合，并从这一视角出发，对高中生的跨学科能力进行了全面的多维度测评和精确的诊断。为了更好地评估学生的能力，陈教授构建了一个测评框架，该框架涵盖了信息转译、反思评价、知识整合和问题解决等四个关键要素。在实证研究中，选择了291名高中生作为样本，并发现这些学生的跨学

科能力普遍较高，且性别和年级差异并不显著。这项研究不仅有助于教师更精准地指导学生，同时也为学生提供了针对性地反馈，有助于他们进一步提升跨学科能力<sup>[32]</sup>。

国内外学者对跨学科能力的理解各异，但其核心基本围绕“互译”、“联系”、“迁移”、“整合”和“创新”等关键点，为本研究构建跨学科能力框架提供了丰富的理论支撑。同时学者对一些已有亦或原创的跨学科综合题进行分析，为本次研究的跨学科试题提供了丰富的材料支撑。有诸多学者已采用跨学科活动，例如“设计和制作简易制氧机”、“电池”、“一次性筷子的利与弊”和“我是火箭推进剂设计师”等主题开展教学实践活动。对学生的跨学科能力进行培养，在测评环节，学者们从科学教育、氧循环、能量平衡等方面进行了深入研究。为本文的测评奠定基础。

### 1.2.3 化学与物理跨学科研究

#### (1) 国外研究现状

2012年，Copp Newton H所开发的综合科学课程序列，旨在融合生物学、化学与物理学的精髓，为有志于未来投身科学专业的学生奠定坚实的跨学科基础。加速综合科学序列(AISS)由资深生物学家、物理学家和化学家联手执教，面向有志于主修自然科学领域的一年级本科生。该序列的课程内容以统一的主题框架进行组织，确保学生在学习的过程中能够形成连贯且深入的科学认知。在教学过程中，学生将在同一课堂内实现从一种活动到另一种活动的自然过渡，而课程安排亦巧妙融合了多种教学方法，旨在全面提升学生的科学素养与综合能力<sup>[33]</sup>。

2013年，Benjamin W. Dreyfus提供了生物学专业本科生物物理入门课程的定性数据，该课程旨在增强跨学科连贯性，包含ATP等主题。学生们在数据中尝试解决ATP中磷酸键断裂时释放能量与断裂键需能量输入之间的明显矛盾。研究发现，学生对各科学学科如何定义兴趣系统的看法会影响其跨学科主题的推理方式。这对跨学科教育的愿景产生影响，该愿景既尊重学科观点，又通过展示一致性将其纳入互动中。尽管生物和化学教育文献中研究了学生对ATP和化学键性质的矛盾看法，但这些主题在物理入门课程中鲜少出现<sup>[34]</sup>。

2024年，Wilcox Jesse展示了我们如何利用一节高中化学课堂上关于反应(HS-PS1-5)的浓度和温度来与消化系统、烹饪和烘焙以及细胞呼吸(HS-LS1-7)建立联系，帮助学生看到学科之间的联系，找到学科核心思想<sup>[35]</sup>。

#### (2) 国内研究现状

在中国知网上以“跨学科”和“物理”和“化学”为关键词进行文献检索，总共有

244 篇，如图 1-4 所示，根据研究内容可分析，绝大多数研究内容为跨学科教学研究。

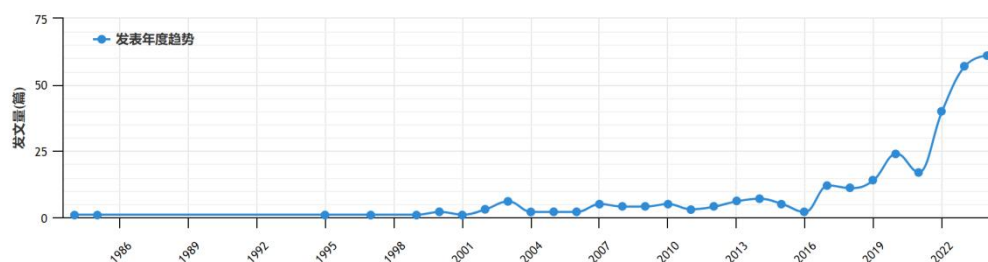


图 1-4 “跨学科”与“物理”和“化学”研究的发文量趋势图

2020 年，李燕磊通过深入研究化学与物理两大学科的交叉领域，精心设计了部分典型教学片段，成功探索出高中化学跨学科教学的新思路。该思路强调在化学教学中融入其他学科知识，进行对比学习，并促进知识迁移。为进一步推广和应用这一教学理念，提出建议：建议不同学科的教师相互听课、评课，以更全面地理解和评价教学过程，共同提升教学质量。建议组建相同学科教师组成的教研小组，共同进行教学设计。建议学校积极开展校本课程，为学生提供更多跨学科学习的机会<sup>[36]</sup>。

2020 年，李文智融合化学领域的“电离平衡”原理与生物学中“pH”稳态的知识，构建了一则名为“平衡：知识的统一”的教学案例。这一案例的主要目的是帮助学生建立不同学科间知识的桥梁，促进他们的跨学科思考。此外，他还运用化学的盖斯定律以及物理学的焦耳实验、内能概念、能量守恒定律和动量守恒定律，精心打造了另一则主题为“平衡：思维的统一”的教学案例。研究结果显示，这种跨学科的教学方法能有效帮助学生深化对知识的理解，并提升他们的课堂参与积极性。此种教学方式不仅有利于全面培养学生的综合素养，还能显著提高他们的学习效果<sup>[37]</sup>。

2023 年，叶鸣扬分析了跨学科实践对教师教学的要求，以及学生自义务教育阶段起常面临的独立学科教学和缺乏学科融合意识的问题。根据相关问题研发了一门名为“溶于于水，融理于化”的跨学科实践课程，该课程将物理与化学两大领域有机地结合起来。同时，他对其他跨学科实践案例进行了深入的反思和分析，旨在寻找学科间融合的通用策略<sup>[38]</sup>。

2023 年，潘书朋以物理化学跨学科融合发展为范例，精心设计了三个实践融合创新实验。这三个实验聚焦于“关联整合、综合应用、创意设计”三大层面，旨在培养学生的物理学科核心素养。为确保实验效果的科学性，他采用了“过程性评价”和“结论性评价”两种评价方式，构建了一套以素养为导向的综合评价体系。这套体系不仅实现了物理、化学学科实验思想和思维方法的深度融合，还提高了学生解决复杂实际问题的能力<sup>[39]</sup>。

#### 1.2.4 相关研究评述

通过对已有的 Rasch 模型研究的分析发现, Rasch 模型非常适用于能力的测评研究, 已经十分成熟, 尤其是对于跨学科能力这样的复杂能力, 国内外都会用 Rasch 模型去测评和检验学生的能力发展情况并且取得了很好的研究成果, 为本研究的测评打下了基础。

经过对跨学科能力研究的深入剖析, 发现相关文献的数量相对少, 其研究内容主要集中在教学研究和能力的培养上, 这两者的研究已经较为成熟, 且取得了一定的成果。然而相对而言跨学科能力测评的研究则显得凤毛麟角。有学者通过整个科学教育视角进行测评, 有学者用“氧循环”、“能量”和“平衡”等跨学科概念角度进行测评, 也有学者通过生物与化学两个学科的整体视角进行测评。但是没有学者从物理与化学两个学科的整体视角进行跨学科能力的测评。再通过对已有的测评研究方法进行分析, 其研究路径通常是基于概念和主题再向学科领域延伸, 这种做法虽然能利用学生熟知的概念和实际生活经验提高其兴趣, 但链接的知识点相对浅显, 难以深入测评学生的深层理解。特别值得注意的是, 现有的研究测评工具大多数都停留在理解层面, 而很少涉及学科之间不同概念的联系与运算, 而运算是理科的共同基础, 因此这一环节至关重要。同时, 目前的研究很少从物理和化学整体的角度进行跨学科测评。基于上述分析, 本研究将运用 Rasch 模型测评规律, 从化学和物理两个学科的视角出发, 并在此基础上加入理科概念之间的相互运算, 开展跨学科测评研究。

### 1.3 研究目的和意义

#### 1.3.1 研究目的

- (1) 根据相关文献界定化学与物理的跨学科能力的内涵。
- (2) 确定化学与物理的跨学科能力组成要素和各要素的水平划分, 从而构建测评框架。
- (3) 根据测评框架和初高中物理和化学学科知识点编制可测量的化学与物理跨学科能力的测评工具。
- (4) 对被试的化学与物理的跨学科能力进行整体分析, 探究被试的跨学科能力的整体表现, 同时对学生跨学科能力的培养提出意见。

#### 1.3.2 研究意义

##### (1) 丰富跨学科能力内涵

在跨学科领域, 对于能力的内涵理解, 学界存在多种观点。本研究旨在通过深入探讨和研究后, 对跨学科能力的内涵要素进行明确界定, 同时对各要素的水平进行科学的划分, 以推动该领域理论体系的进一步完善。

##### (2) 提供跨学科教学后的评价工具

在实施跨学科教学后, 教师可利用测评工具对教学成果进行检验。基于 Rasch 模型所编制的工具能够将学生内隐能力转化为外显表达, 从而更准确地反映教学效果。这为

一线教师提供了新的评价方式，进一步提升了教学效果评估的准确性。

### (3) 促进学生跨学科能力的发展

通过软件工具对数据进行分析，教师可以更加精确地识别学生的能力短板，以便制定更符合学生需求的个性化教学方案。这种方法不仅有助于提高教学质量，还能帮助学生更清楚地了解自己的学习状况，进一步发挥自我监控和调节的作用，从而促进学生个人能力的全面发展。

## 1.4 研究方法

### (1) 文献分析法

在中国知网上以“Rasch 模型”、“跨学科能力”、“化学与物理跨学科”等为关键词进行文献检索，对搜集到的文献进行梳理、整合，分析总结 Rasch 模型、跨学科能力的国内外研究现状，基于现有的研究成果和研究不足，寻找本课题研究的突破口和着力点。

### (2) 教育测量法

利用测试卷对目标群体进行了纸笔测验。首先构建了能力水平框架，该框架基于跨学科的综合解题过程以及相关文献的深入分析。随后根据该框架精心设计和挑选了测试题目，以确保能够全面评估受试者的能力水平。在选择测试样本时，充分考虑了研究目的，以确保数据的准确性和可靠性，选择了合适的学生样本进行测试，并确保了答题纸的完整回收。最后，对收集到的数据进行了详细的统计、分析和处理，得出研究结果和结论。

### (3) 统计分析法

在获取充足的实证数据之后，将进行数据的有序编码、精确录入和深入分析。为确保评估的科学性与公正性，本研究将采用 Rasch 模型作为核心评估工具。同时，还将利用 Microsoft Excel 2020 以及 Winsteps 3.72.3 等先进软件，对数据进行细致地处理与分析。分析的重点将聚焦于评估工具的信效度检验，全面揭示学生的整体表现特征，并深入探究学生在不同能力要素上的水平差异。

## 1.5 研究思路

本文通过分析已有研究，通过理论与实践相结合，根据相关研究构建跨学科能力框架模型，并且结合物理与化学学科知识点开发并优化跨学科能力测评工具测评学生的跨学科能力，通过分析学生跨学科能力现状并提出培养建议，研究思路如图 1-5 所示。



图 1-5 研究思路图

## 2 概念界定与理论基础

### 2.1 跨学科能力的界定

跨学科融合的概念起源于 20 世纪 70 年代，并在国际论坛上首次由经济合作与发展组织（OECD）提出。这一理念有着宽广的含义，它覆盖了从思维交流到实践层面的各个方面，包括但不限于组织原则、研究策略、流程管理、认知科学、术语系统、信息处理，以及教育和研究机构之间的跨领域合作<sup>[40]</sup>。跨学科融合的关键在于促进认知层面的整合，并且鼓励不同领域的知识相互融通和重组，从而加深和创新知识体系。在过去的数十年里，以创新为目标的多学科研究方法日益受到学术界和社会的重视。关于跨学科知识的界定，学界存在多元看法。经过与 12 位多学科研究领域专家的深度对话，John D. Aram 总结出两种不同的定义。有些学者强调，跨学科知识产生于不同领域内知识的互相整合，从而开创出新的知识结构。他们认为跨学科知识的综合性是其特点之一，它跨越了传统学科的界限。另外一些学者则有所区别，他们认为多学科知识并非单纯的整合产物，而是基于不同领域间的相互作用，通过知识的迁移或结合，诞生出新的视野或理论。这一观点看重学科间知识的互动，认为跨学科知识是建立在单个学科基础上的延伸。这两条思路虽然各有侧重，却共同凸显了跨领域知识的特质与形成过程<sup>[41]</sup>。

Jones 和 Merrit 强调，跨学科能力是一种独特的能力，它能够将不同学科的知识进行有机融合。每个学科都有其独特的贡献和视角，为知识的宝库增添了丰富的色彩。他们认为，跨学科不仅仅是多个单一学科的简单相加，而是对各个学科领域知识的深入理解和综合应用<sup>[42]</sup>。

基于知识整合层面和应用层面的角度，Guerrero 和 Reiss 提出了“理论跨学科”和“方法论跨学科”<sup>[43]</sup>。“理论跨学科”是一种认识论形式，旨在建立一个概念框架，整合命题和概念，链接跨学科的主题，并整合模型和类比之间的连续性。这种形式的跨学科可以使学科间的知识得以深入交流和互补，还能够从多个角度分析并解决问题<sup>[44]</sup>。而“方法论跨学科”是指在跨学科研究中，研究者从其他学科借用方法来解决本学科的问题。这种方法使得研究者能够充分利用其他学科的研究成果，提高研究的质量和效率，同时也在某种程度上推动了各学科之间的交流和融合。方法论跨学科性的优点在于，它可以帮助研究者从不同的角度和方法来解决问题，从而得到更全面和准确的研究结果。同时，这种方法也有助于促进各学科之间的合作和交流，加速学科间的融合和整合。然而，方法论跨学科性也存在一些挑战，例如不同学科之间的方法和理论的差异可能导致方法的不可靠性和效率低下等问题<sup>[45]</sup>。

经过文献的深入剖析，跨学科的概念不仅涉及不同学科间理念的相互渗透，同时也包括特定学科领域的整合。因此，跨学科的范围具有一定的伸缩性，能够适应各种应用

场景。鉴于当前学科的发展走向以及高中生的课程学习情况，本研究选择自然科学领域的化学与物理作为研究内容，对高中生进行跨学科能力的评估。此种能力被定义为化学与物理的跨学科能力。

本研究基于化学与物理学科的视角，将跨学科能力界定为一种综合性的能力。这种能力表现为在面对复杂的科学挑战时，学生能够运用化学、物理学科的知识 and 思维方式，全面而整体地把握问题。在此过程中，学生需要积极投入，包括识别和分析情境、明确各学科的核心概念，以及深入反思信息和观点等。同时，学生还需要有效整合跨学科的理论知识框架和研究方法，最终提出针对问题的解决方案或合理的解释。这种能力的培养对于提高学生的科学素养和应对复杂问题的能力具有重要意义。

## 2.2 理论基础

### 2.2.1 学习迁移理论

学习迁移理论和跨学科能力之间存在着紧密的关系。学习迁移理论指的是通过学习一种知识或技能来帮助学习其他知识或技能的能力<sup>[46]</sup>。跨学科能力最本质的内核就是迁移，有知识的迁移和方法的迁移，因为在学习过程中，不同的学科领域之间往往存在着相似性和关联性。比如说，在学习数学知识的过程中，需要运用到逻辑思维和分析问题的能力，在学习物理知识的过程中，需要运用到数学和几何知识等等。因此，具备跨学科能力的学生更能够将不同学科领域的知识和技能进行整合和应用，从而更加有效地实现学习迁移。同时，学习迁移和跨学科能力也是相辅相成的。通过不断地进行跨学科学习，在不同学科领域中积累知识和技能，可以更加有效地实现学习迁移。通过学习迁移，能够更加深入地理解和掌握不同学科领域的知识和技能，从而更好地发挥跨学科能力的作用。而在本研究的测评工具中，学生需要将所学习的物理、化学知识相互迁移并且综合应用才能达到解决问题的目的。因此，学习迁移理论和跨学科能力之间是密不可分的关系。

### 2.2.2 多元智力理论

在学术界，多元智力理论与跨学科能力的培养之间的关系备受关注。按照多元智力理论的观点，人的智力是由多种不同的智力要素构成的，这些要素包括逻辑数学智力、言语语言智力、空间视觉智力、身体动态智力、音乐节奏智力、人际交往智力和自我认知智力等。这一理论认为每个人的智力都是独一无二的，而不同智力能力的组合也会对个体在不同领域的表现和发展产生影响<sup>[47]</sup>。在培养跨学科能力的过程中，需要运用多种智力能力，而多元智力理论为此提供了一种理论基础。这一理论有助于人们更好地理解和发展不同的智力能力，因此在实践教学环节中，应充分融合其他学科的知识、方法和思想。这不仅能为学生提供更广阔的发散思维空间，还能调动他们多种智能的应用，使学习更具多元性、趣味性和高效性。而在本研究中，学生需要用到逻辑数学计算能力，

因为题目都是理想化模型，需要学生在做题时需要用到空间视觉想象智力，还有逻辑推理智力等许多智力，才能达到解决问题的目的。

### 2.2.3 建构主义学习理论

皮亚杰的建构主义学习理论主张，学习者并不只是被动地接受知识，而是通过与周围环境的互动，逐渐构建和重塑对知识的理解。在这个理论框架下，知识不是预先存在的，而是由个体主动构建的，这个构建过程是学习者与环境相互作用的结果<sup>[48]</sup>。在自然科学领域，化学、物理和生物三者间存在着密切的内在联系。因此，当学生在构建化学知识体系时，教师要创设有利于学生将各个学科联系起来的学习情境，积极引导调动已掌握的物理和生物知识，主动将其与化学知识相融合。这样不仅能加深学生对化学知识点的理解，还有助于建立起学科间的有机联系，形成完整、统一的知识结构，避免各学科之间的孤立与隔阂。

### 2.2.4 项目反应理论

#### (1) Rasch 模型的数学表达式

Georg Rasch 首次提出的 Rasch 模型是项目反应理论的关键部分。它通过数学模型来探究参与者的能力水平与试题难度的深层次联系。为了保证模型的精准度和信度，通过实际的测试数据对其进行了拟合度检验。此外，该模型采用 logistic 函数来进行详细的数学描述，详见图 2-1<sup>[49]</sup>：

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

图 2-1 Rasch 模型的数学表达式

在 Rasch 模型的框架下，用于预测一个参与者正确解答特定问题  $i$  的几率，记为  $P_i(\theta)$ ，是通过一个特定的公式进行计算的。其中， $\theta$  表示参与者的内在能力水平，而  $b$  则是衡量问题难度的参数。值得注意的是，在 Rasch 模型中，每一题的特征只由单一参数，即题目的难易程度，来定义。这个模型的核心在于简化问题与参与者能力之间的关系，通过比较题目难度与个人能力的高低来计算答题成功的概率。

#### (2) Rasch 模型的主要特点

##### ① 被试与试题共用一把尺测量

Rasch 模型通过利用对数转换的方法，实现了把个体和题目在同一维度标尺上的评估。此模型通过一个连续单维度的框架，使我们能够直接比较不同个体、不同题目之间，以及个体与题目之间的相关位置。这种做法区别于传统的测量技术，它的独特之处和优势在众多实际应用场景中得到了体现<sup>[50]</sup>。

##### ② 数据的线性

Rasch 模型的核心优势在于，它能够把本质上非线性的信息转化为具有等间距属性的对数几率尺度（logit scale）信息，这样的转变让对特定特性的客观评估成为现实。这个模型的独特之处在于，它被一些研究者视为将序数级别的数据有效转化成连续线性形式的唯一方式<sup>[51]</sup>。

### ③参数的不变性

当进行能力或成绩的评价时，传统的测试手段常面临一些限制，在对个人不同考试成绩的对比和预期成绩的预测方面尤为明显。然而，Rasch 模型通过确立了独立的个体能力参数和试题难度参数，突破了这些限制。这种独立性符合在量度方面追求的参数分离原则，允许更精确的个体能力评估。据研究表明，这种方法因其参数的独立性，相比传统评价体系，在预测个体表现方面显示出更强的准确度<sup>[52]</sup>。

### （3）Rasch 的主要参数<sup>[53]</sup>

表 2-1 测评工具质量指标参数含义及取值范围

质量参数		含义	参数指标
信度	信度	测验工具的一致性、稳定性	可接受范围为 0-1, 大于 0.8 较好
	分离度	被试能力和测验难度在测量变量上的离散程度, 为了区分不同被试的能力, 分离度越大, 区分效果越好	大于 2 较好
	误差	测量值与期望值的差异, 反映测量的精确程度	接近 0 较好
效度	项目-拟合	实际与理论反应模型之间的拟合度	MNSQ (Infit, Outfit) 可接受范围为 0.5-1.5, 理想值为 1。 ZSTD (Infit, Outfit) 可接受范围为 -2-2, 理想值为 0
	单维性	测量的是单一特质	-0.4-0.4 之间较好
	点测量相关	项目与总量表之间的一致性	可接受范围为 0-1, 理想值为 1

### 3 化学与物理跨学科能力测评框架的建构

#### 3.1 化学与物理跨学科能力的组成要素

##### 3.1.1 文献中的跨学科能力模型

综合考虑了跨学科能力的多样性和复杂性,许多专家学者已经从多个角度对跨学科能力进行了详细的分析。他们的工作不仅揭示了跨学科工作的各个层面,也为教育者提供了可执行的方案,以便于更精准地进行跨学科教育活动的指导和评价。这些方案中包含了一系列能力指标和培训方法,旨在评估和提升个体在跨学科场景中的表现。

在教育界中享有盛名的 Klein 先生强调,跨学科的学习是一个动态的进程,其中信息的搜寻与知识的应用是紧密相关的,并且涉及到若干关键过程,包括提出问题、解决问题、做出决策以及进行深度的批评性思考。要想在跨学科领域取得成功,学习者必须掌握一系列重要的技能,这些技能对于跨学科学习的成功至关重要。这些关键技能主要包括:寻找复杂问题的深刻见解的能力;能够从多个渠道获取、整合知识和观点的能力;利用比较和对照来发现模式和关系的能力;以及能够建立全局观以获得更深入理解的能力。通过这样的描述,可以看出跨学科学习不仅要求对信息的有效管理,还需要能够批判性地分析和思考,以构建对复杂情况的全面理解<sup>[54]</sup>。

在探讨学科间合作的核心,许多研究者指出了知识融合的重要性。在 2006 年, Linn 与 Eylon 率先提出了对知识融合的理解,并将其分成四个步骤。最初步是回顾既有见解,随后是引入新的看法或方法以丰富这些见解<sup>[55]</sup>。接着,是对众多见解进行鉴别和评估,最终是将这些见解进行有系统的梳理和总结。之后, Linn 及其团队对这一理论框架进行了仔细的修正,并最终形成了一个在教育研究界得到广泛认可的知识融合框架。基于此框架,他们还提出了用于评价知识融合程度的五个级别的标准。具体如表 3-1 所示<sup>[56]</sup>。

表 3-1 知识整合框架的五个水平

回忆已有观点	帮助学生唤醒和激活他们已有的观点、背景和经验,为他们在 Learning 环境中构建新旧思想的桥梁做好准备。
添加新的观点	通过组织有针对性的学习活动,引导学生发掘新的观点,深化对已有观念的认识,并建立其与新思想的有机联系,促进学生思维的拓展与整合。
区分观点	学生需对生产性观点与非生产性观点进行深入思考和区分,以确保形成的观念与科学相关且具有规范性。
反思观点	针对相互冲突的既有观点,需要进行系统地梳理和排序,深入剖析这些观点之间的内在联系,从而建立起对主题的全面而连贯的理解。
制定标准	为确保科学理解的一致性和持久性,需要明确观念间可能存在的各

---

种关系，并为其设定相应的标准。

---

在知识整合模式下，Shen 等人明确指出其三个特殊步骤：互译、适应与创新。知识整合理论认为：学生的思考经常呈现为碎片化形态，故而科学教育的关键在于助力学生整合不同观点，形成连贯的阐释，从而解决复杂的科学问题。以植物的光合作用为例，从物理学角度，关注光能如何转化为化学能；从化学角度，研究具体的化学反应过程；而从生物学角度，探索植物内部结构与功能在光合作用中的角色。对光合作用的全面理解，需要整合这三个学科的知识。

(1) 互译：当前教育体系以分科教学为主，科学类学科虽然共享一些基础通用术语，但每个学科领域都有其独特的专业名词。鉴于各学科领域专有名词的存在，跨学科的知识整合需要进行术语的统一或翻译。值得注意的是，在某些情况下，不同学科对同一术语的定义可能存在差异，而相同或相似的术语在不同的学科中也可能具有不同的含义。因此，学生在学习过程中需要具备在不同学科间转换术语的能力，以便更好地理解 and 建立各学科之间的联系。

(2) 适应：要求学生能够将一个领域的理论或概念应用到其他学科，从而达到跨学科的深入结合，以理解和解决该学科领域的问题。而实现这种适应的关键，在于学习者能否识别不同学科所共有的核心结构。这种不是停留在表面，而是深入到学科的本质。在现实生活中，学生需要不断地将所学的知识从一个情境适应到另一个情境，以适应不断变化的环境和问题。

(3) 创新：学习者在掌握某一学科的概念、原理和方法后，能够灵活地将其应用到其他学科中，通过创新的方式实现概念或物质层面的转变，从而产生新知识或创造新产品。这种跨学科的实践和探索，要求学习者具备扎实的学科基础和严谨的思维方式，以实现不同领域间的相互融合与提升。

在跨学科理解的认知过程中，互译、适应和创新三者相互交织，并非各自独立。准确互译不同学科的专业术语是实现知识适应的先决条件，而知识的创新则是在熟练掌握互译和适应的基础上得以实现<sup>[15]</sup>。

经仔细审视相关文献和研究，已经确认多个框架或理论模型定位了跨领域技能这一概念，并广受认可。这些建构不单展示了跨学科技能的逐步认知进展，同时包含了如何将认知技能与其他综合性科学能力相融合的复合视角。在此研究内，跨学科技能被定义为一个包含横向维度，即不同学科知识的转移与结合，以及纵向维度，指更深层的信息整合、转移和推理能力的综合构成。如此全面与深远的界定，为未来跨学科能力的探究及其应用提供了扎实的理论支撑。

### 3.1.2 化学与物理跨学科能力组成要素

本项研究针对跨学科能力在化学与物理领域的应用进行了深入探讨。研究借鉴了现有的文献，并对多学科问题解决的步骤进行了详细分析。研究成果明确了跨学科能力的

四个主要成分。基于这些发现，研究者们设计了一个评价体系，以便对学生的跨学科能力进行合理的评估。考虑到高中学生的学习情况与其认知水平，本研究专注于化学和物理两个学科领域，旨在为跨学科技能的评估和诊断提供一个量化的标准。

### （1）信息的提取与转译

在处理跨学科问题时，常常会面临大量复杂且多样的背景信息，这些信息不仅数量庞大，而且科学准确性难以确定，甚至可能包含某种引导性。因此，为了有效地解决这些问题，学生必须具备从复杂情境中提炼关键信息并深入理解问题本质的能力。在此背景下，本研究提出了跨学科能力的首个核心要素——“信息的提取与转译”。这一要素主要关注学生在面对跨学科问题时，能够从复杂的情境中准确提取关键信息，并能够根据已有知识结构识别并解释不同学科中同一概念的不同解释及专业术语，例如化学中的焓变与物理中的热值，两者都代表了一个物质反应产生的热效应，但是它们在两个学科的定义和数量关系有所不同，学生就要结合自己对这两个概念的理解，将两个概念合理地转译。这种能力不仅有助于学生更好地理解和应用不同学科的知识，而且还能更有效地解决这些复杂的跨学科问题。

### （2）知识的整合与联系

在处理跨学科问题时，一旦学生准确地识别出问题中的核心概念，他们需要主动将这些概念整合到跨学科的框架中。这种整合不仅增强了单一学科的解释能力，而且形成了新的知识结构，为深入理解问题的本质提供了独特的视角<sup>[57]</sup>。考虑到现实世界的复杂性，单一学科视角往往显得捉襟见肘，而跨学科的整合方式则能让学生更全面地阐述自己的见解，进而解决跨学科问题并推动认知的发展。本研究着重强调知识整合在学生跨学科能力发展中的重要性。当面临跨学科问题时，学生的首要任务是在复杂的情境中精确捕捉问题的关键信息。接下来，需要将这些信息与他们已有的化学和物理知识相结合。这种结合不仅促进了不同学科知识的相互关联，还培养了学生独特的跨学科思维方式，使得化学和物理知识得以相互启发和深化<sup>[30]</sup>。

本研究将知识的整合与联系视为构成学生跨学科能力的关键要素之一。在面对跨学科问题时，学生需要具备在复杂情境中精准定位问题信息，并对其进行有效提取、转译和批判性审视的能力。进一步地，学生需要将分散的化学和物理知识进行有机整合，建立不同学科之间的内在联系，从而形成一种跨学科的思维方式。这种思维方式有助于学生在解决实际问题时，能够更加全面、系统地思考问题，提高其综合运用知识的能力。例如学生在学习原电池和电解池的时候，可以迁移到物理里的电路、电流、电压和电阻等相关概念，进行联系和整合。

### （3）概念的融合与量化

在理科学习中，计算能力无疑占据着举足轻重的地位。在化学与物理的知识当中，很多概念都有量化的关系，而无论是物理还是化学，学生在应对考试时常常因无法妥善

处理计算问题而倍感困扰。同时值得注意的是，相较于定性研究，定量研究方法具有更强的说服力，这是必须正视的事实。当学生能够将不同学科的知识融会贯通时，他们将更深入地领悟到相通概念在不同学科中的具体表现。然而，需要强调的是，不同学科间共有的概念在数学定义上往往存在显著差异。因此，为了更准确地解决问题，学生必须清晰地理解其中的数量关系。

例如在物理学中，热值与化学中的反应热虽然在本质上具有相似性，然而，由于它们单位定义的不同，数量关系并不直观。具体到化学反应与能量之间的关系，例如化学反应所产生的能量可以为各种运动提供动力。然而，这些数量关系在教材中并未得到明确地体现。因此，学生需要运用已掌握的知识，探究它们之间的数量关系，并通过计算来解决实际问题。

本研究将概念的融合与量化作为衡量学生跨学科能力的第三个要素。这意味着，在学生对不同学科的概念进行融合后，他们应能展示两个相通概念之间的数量关系。通过计算，他们可以更直观地表达和理解这些概念间的关系，进而提高解决问题的效率。

#### (4) 问题的解决

面临科学上的复杂问题，单靠积累的科学知识往往难以完全应对挑战。需要的不仅是对已有知识的识别、定位和融合，更关键的是具备将这些知识跨领域应用的能力，以及将思维方式从一个领域转移到另一个领域的灵活性，这些构成了跨学科素养的精髓<sup>[58]</sup>。研究表明，公众在使用科学方法来解决日常问题时的总体意识和能力尚有待提升。培育跨领域的技能的根本目的是使人们在遭遇新的日常挑战时，可以借助于广泛的知识体系和综合的思辨能力进行科学的逻辑推断和实际应用，以便有效处理问题<sup>[59]</sup>。

在科学领域，各个学科之间的边界模糊、交织着众多共有的理念和基础理论模型。当面对需要涉及多个学科知识来解决的问题时，学生们必须以一门学科的理论为出发点，汲取并整合另一学科的关键思想来进行分析和解释。特别是对于那些不归属于任何特定学科的问题，对学生的能力提出了更高的要求：他们必须基于知识整合的能力，采用跨学科思维方式，建立起问题的核心架构，并能够将此应用于新的或未知的场景中以寻找解决方案。这种跨学科能力的展现，不仅显示了不同学科之间知识的转移能力，还反映了各学科基本素质之间的相互影响和转化。

本研究引入了一个至关重要的跨领域学习素质—问题解决能力。通过考察结合化学和物理学的复杂问题，可以看到拥有这种能力的学生如何利用一个领域的基础理论或框架来阐释在另一领域遇到的全新问题。他们不受单一领域限制，能基于跨领域的综合视角，把不同领域的核心知识应用到新的问题场景中，并能够提出恰当的解释方案。比如，在分析化学电池时，他们会从能效、反应产物及其再利用等多方面进行综合评估与优化建议。

根据上述分析，本研究确定了跨学科能力的四个组成要素，分别是“信息的提取与转译”、“知识的整合与联系”、“概念的融合与量化”和“问题的解决”，解决复杂的跨学科问题并非简单的线性认知思维过程，而是涉及多种综合能力的协同作用。在解决这类问题的过程中，信息的提取与转译是基础前提，知识的整合与迁移则是建立各领域间联系的关键环节，同时，定量计算作为一种重要的手段，有助于学生更精确地分析和解决问题，只有当这些要素共同作用时，才能有效地解决跨学科的复杂问题，实现知识的迁移。各个能力组成要素如表 3-2 所示。

表 3-2 跨学科能力组成要素及要素内涵

能力组成要素	要素内涵
信息的提取与转译	在面对错综复杂的跨学科问题时，学生需具备从生疏情境中精准提炼核心信息的能力，并能够从已掌握的知识体系中识别并比较不同学科对同一概念的差异性解读，以专业、严谨的态度解读各学科专业术语。
知识的整合与联系	学生需要将题目中的信息与自身所学的化学和物理知识相互联系起来，进一步建立多种知识点之间的内在联系，从而形成一种跨越化学和物理学科的独特思维方式。
概念的融合和量化	学生经过对不同学科概念的融会贯通，能展示不同学科相通概念之间的数量关系，通过已学的公式进行计算，准确表达两个相通概念之间的数量关系。
问题的解决	学生能够运用某一学科的核心概念或模型，有效地解释另一学科领域的新情境。他们具备学科交叉思维，能够超越单一学科的限制，将不同学科的知识整合起来，将所学的知识从一个情境迁移到另一个情境，灵活地应用于新的问题情境并提出合理的解释和解决方法。

每一个问题可能包含一种或多种要素能力，因为这些要素不是相互独立的，而是一种递进和包含的关系，所以在后面在题目与能力对应时，选取主要的考察要素为对应。

### 3.2 化学与物理跨学科能力模型的建构

本研究的核心目的是开发一种评估系统，专门用以确定并分析高中学生的多学科综合能力。鉴于学生之间在能力和成长轨迹上的差异，评估这些能力就需要一个更加客观和精确的办法。因此，本工作深入研究了多学科能力的关键要素，并将它们分类，从而把原本难以捉摸的能力概念具体化，并形成可行的评估架构。这个初步构思的框架对于后续开发详细的评估工具极为关键，它能有效地帮助评定学生在多个学科间的综合应用能力。

### 3.2.1 信息的提取与转译的水平建构

在处理问题时，首要步骤是确保对题目信息的准确识别和理解，深入挖掘题目的内涵。同时，需要将题目中的信息与学生脑中的知识体系和生活经验进行有效地整合。为了更准确地把握题目的含义，有时需要将日常生活中的通俗表述转化为科学术语。在后续的研究中，将从信息的隐蔽程度和与科学概念的关联准确性两个方面对这一要素进行分类和探讨。具体表现见表 3-3。

表 3-3 信息的提取与转译发展水平及其表现期望

发展水平	期望
水平 1	当学生面对跨学科问题时，能够从隐秘程度低，信息明确的情境中提取出部分关键信息。
水平 2	当学生面对跨学科问题时，能够从有较多隐秘信息的情境中找出全部关键信息，并唤起已学的相关概念。

### 3.2.2 知识的整合与联系的水平构建

在学生的认知结构中，原有的科学概念和生活经验已被充分唤醒，并且已经注入了大量的经过处理的“新知识”。然而，这些新知识与旧知识之间，以及不同学科之间的内在联系尚未稳固建立。为了能够切实解决跨学科的问题，一个完整且连贯的知识网络，以及一个跨学科的整合视角，是不可或缺的。本研究以知识整合理论为基石，依据知识间的连接程度来评估知识的整合与联系的水平，进而构建相应的评估指标。具体表现见表 3-4。

表 3-4 知识的整合与联系发展水平及其表现期望

发展水平	期望
水平 1	能够理解题目信息，与已有的各个学科知识建立初步联系，用简单的跨学科视角理解问题。
水平 2	能够把来自不同学科的知识点熟练地连结起来，并且把这些知识整理排序，形成结构化、系统化的知识网，形成跨学科的连贯理解，用跨学科的眼光来看待问题。

### 3.2.3 概念的融合与量化的水平建构

数量关系的计算是理科学习的重要基础，它对于深入理解学科本质至关重要。只有在定性分析的基础上进行定量分析，才能更全面地理解学科本质，并有效地解决问题。本研究将学科概念之间的链接复杂性作为划分这一素养的标准。具体表现见表 3-5。

表 3-5 概念的融合与量化发展水平及其表现期望

发展水平	期望
水平 1	能够理解题目中学科某些概念之间直接的数量关系，利用公式进行简单的

	推理计算。
水平 2	在准确定位题目中大量的学科概念后，能调动已学知识中关于各个概念的相关公式，并且建立公式之间的关系，进行计算。

### 3.3.4 问题的解决的水平建构

在面对跨领域问题时，学生们不只需要确立扎实的科学知识基础，还必须有能力将现有的知识架构与解决方案灵活运用于新的场合中。解决问题是认知能力的高级形式，而解决跨领域问题不同于其他类型问题解决的是其多变的情形及合并不同知识所需处理的复杂度。通常在教学过程中，学生透过反复练习特定种类的问题来熟悉解答方法，他们已经对如何处理作业和考试中的常规科学问题有了充分的了解，并能有效地联合和运用各种相关知识来构建答案。尽管如此，这种重复性练习可能会限制思维的灵活性；一旦遭遇非典型情形或需要偏离现有思考模式的问题时，学生通常会遇到难题。因而，为了更好的迎接跨领域问题带来的挑战，关键在于拓展学生的思考视野和增强他们的问题处理技巧。

本研究将跨学科能力评估置于与学生平时的练习有所不同的陌生问题情境中，需要学生以化学、物理两个学科为基础，解决跨学科问题。本研究将根据问题情境的熟悉度、知识广度，按照单一学科，多学科的进阶模式进行划分。具体表现见表 3-6。

表 3-6 问题的解决发展水平及其表现期望

发展水平	期望
水平 1	能够从每个学科不同的角度来认识问题，并且能够运用不同学科的知识独自合理地解决一些常见的科学问题。
水平 2	能够以整合的跨学科视角，将不同学科的知识结构整合，迁移到新的、复杂的情景，解决各种陌生的问题。

## 4 测评工具的设计和优化

### 4.1 测评工具的开发流程

本研究通过设计试题，根据学生纸笔测验的作答情况，按照评分细则进行赋分，利用 Excel 收集数据，在用 Winsteps 软件进行分析学生跨学科能力的水平情况，具体流程如表 4-1。

表 4-1 测评工具的开发流程

流程	具体实施步骤
能力框架的构建	根据文献研究和分析的结果，对当前的能力框架进行了梳理，并构建了跨学科能力的能力测评框架。该框架包括四个要素，每个要素都有不同的水平划分。具体细节请参考上文。
试题编制	查找相关文献和中、高考题，根据化学与物理学科知识的实际情况，设计与能力框架中各能力要素相匹配的试题。
评分标准	根据能力框架，预测学生可能回答，初步设计评分标准。
试测和修订	请一线老师（中学物理和化学老师）和化学教学论的研究生同学先做一遍试题，根据他们的意见进行修改，再让一些中学生进行初步测试。再根据学生的回答情况进行修改，形成最终试题。
测评实施	进行大规模学生纸笔测验。
质量分析	依据 Rasch 模型，运用 Winsteps 软件对被试和测评工具进行信、效度检验。

### 4.2 测评工具的开发与设计

为确保试题内容的准确性和合理性，绝大多数试题都是文献中和中、高考出现过的试题，加上自己对跨学科能力理解后进行编写，同时与一线教师积极讨论并修改，试题知识内容全来自初、高中的物理、化学内容，题型全为主观题，避免了选择题的猜测的可能。详细知识点对应物理和化学内容如表 4-2（高中教材使用人教版，初中教材使用科粤版）。

表 4-2 初始测试题与知识点对应表

题目	化学知识点（对应教材大致位置）	物理学科知识点（对应教材大致位置）
1	燃烧的条件 （初中化学九年级上册第三章第三节）	摩擦做功 （初中物理九年级上册第十一章第一节）

2	燃烧的条件 (初中化学九年级上册第三章第三节)	熔沸点 (初中物理八年级上册第四章第二节)
3	碳酸钙和稀盐酸反应 (初中化学九年级上册第五章第三节)	浮力 (初中物理八年级下册第九章第一节)
4	铁的生锈 (初中化学九年级下册第六章第四节)	浮力 (初中物理八年级下册第九章第一节)
5	离子浓度与导电性 (化学必修第一册第一章第二节) 浓硫酸的性质 (化学必修第二册第五章第一节)	欧姆定律 (初中物理九年级上册第十四章第二节)
6	质量分数 (初中化学九年级下册第七章第三节)	浮力 (初中物理八年级下册第九章第一节) 杠杆平衡 (初中物理八年级下册第六章第五节)
7	焓变、燃烧热 (化学选择性必修一第一章第一节) 侯氏制碱法 (化学必修第一册第二章第一节)	热值 (初中物理九年级上册第十二章第二节)
8	碳酸钠与碳酸氢钠的性质 (化学必修第一册第二章第一节)、溶液浓度与密度 (化学必修第一册第二章第三节)	浮力 (初中物理八年级下册第九章第一节)
9	原电池、电解池 (化学选择性必修一第四章第二节)	磁场 (物理必修第三册第十三章第一节) 安培力、感应电动势、磁通量 (物理必修第三册第十三章第二节) 受力分析 (物理必修第一册第三章第四节) 电量与电子 (物理必修第三册第十一章第一节)
10	气体摩尔体积 (化学必修第一册第二章第三节) 电解水 (初中化学九年级上册第四章第二节)	能量的转化 (物理必修第二册第八章第四节) 电量与电子 (物理必修第三册第十一章第一节)

		能量守恒定律 (物理必修第二册第八章第四节) 影响导线电阻的因素 (物理必修第三册第十一章第二节)
11	铁与稀硫酸反应 (初中化学九年级下册第六章第二节)	压强 (初中物理八年级下册第八章第一节)
12	葡萄糖的氧化 (初中化学九年级下册第九章第四节)	机械效率 (初中物理九年级上册第十一章第三节) 能量守恒 (物理必修第二册第八章第四节) 加速度 (物理必修第一册第一章第四节)
13	电子转移 (化学必修第一册第一章第三节) 离子电荷 (化学必修第一册第一章第二节) 电镀 (化学选择性必修一第四章第二节)	电量 (物理必修第三册第十一章第一节)
14	电解饱和食盐水 (化学选择性必修一第四章第二节) 物质的量、气体摩尔体积 (化学必修第一册第二章第三节)氯气的性质 (化学必修第一册第二章第三节)	机械效率 (初中物理九年级上册第十一章第三节) 能量守恒 (物理必修第二册第八章第四节)

根据当前高中科学课程的设置情况，本研究选取了高中学段（高三年级）的化学、物理选考考生作为样本对象。在测试卷中，考察的项目所涉及的化学和物理等科学知识的难度均未超过高考的知识点要求。这些知识点主要来自初中阶段和高中必修阶段的课程，并且只有少量的知识点属于选择性必修课程。因此，这些知识点对于化学、物理选考考生来说是需要掌握的内容。然而，此次试题需要学生通过学科之间的联系进行拓展，其题目的考察形式与学生平时所做的练习题有较大区别，因此对学生的能力有一定的挑战。

根据跨学科能力的框架建构，一共设计了 14 道题目，构成测试卷的初试试题。（具体见附录一）测试卷中每一题的解决过程都需要学生运用到四种要素的跨学科能力，

因此，将每一个问题比作一个项目，每一个项目对应一项跨学科能力（有些题目可能包含多种能力，这类试题以主要测评的能力为主），其中对应情况如表 4-3。

表 4-3 初始测试题与能力水平对应表

能力要素	水平	题目
信息的提取与转译	1	8 (1), 8 (3), 12 (1), 14 (1)
	2	7 (1), 13 (1), 13 (2)
知识的整合与联系	1	1, 2, 3, 4, 6 (1), 7 (2), 9 (1)
	2	5, 6 (2), 8 (4)
概念的融合和量化	1	8 (2), 7 (3)
	2	9 (2), 10 (1), 10 (2), 10 (3), 11, 12 (2), 13 (3), 14 (2)
问题的解决	1	10 (4), 12 (3), 8 (5)
	2	7 (4), 9 (3), 13 (4)

以下在不同要素水平下选取一道题为例进一步说明试题的开发过程，同时对评分标准进行阐述。

(1) 信息的提取与转译能力水平 1 和 2 的试题分析：

第 8 题的第 1 小问：物质的溶解度和该物质饱和溶液的密度之间有什么关系

这里要求学生从化学学科回忆溶解度的相关概念，从物理学科回忆密度，能理解溶解度大，饱和溶液越浓，溶液密度越大就可以做出此题。这道题给的信息十分直接，对溶解度与密度进行直接的信息分析，就可以得出答案。所以此题考察的能力水平为信息的提取与转译水平 1。

①回答出来则为 1 分，

②回答错误得 0 分。

第 7 题的第 1 小问：根据材料可知联合生产的优点是什么？副产品是什么？有何作用？

这道题就需要学生进行材料分析，理解何为联合生产的含义，可以得出联合生产的优点是节约资源，绿色环保，再结合材料后段的信息：石灰窑中产生的富含  $\text{CO}_2$  的窑气通入氨的氯化钠饱和溶液中，40%的  $\text{CO}_2$  最终转化为纯碱。再回忆侯氏制碱法可知副产品为氯化铵可以作为氮肥，相对于水平 1，这题的信息较为模糊，需要学生对题目中的联合生产和侯氏制碱法进行系统的信息分析，才能得出结论，所以此题考察的能力要素为信息的提取与转译水平 2。

①如果学生回答出联合生产的优点（答案合理即可），则为 1 分。

②如果学生也能回答出副产品和用途，则再加 1 分。

③回答错误为 0 分。

(2) 知识的整合与联系能力水平 1 和 2 的试题分析：

第 9 题的第 1 小问：求金属棒刚开始运动时，蓄电池为充电状态还是放电状态？

这道题学生需要联系电磁感应和电解池的相关知识，通过得出感应电动势和蓄电池电动势的比较，可以得出充放电状态。此题虽然有一点计算，但是只涉及物理方面的计算，再通过判断即可，这道题将电磁感应和电解池进行联系整合，所以此题考察能力水平为知识的整合与迁移水平 1。

①如果学生计算出感应电动势，则为 1 分。

②如果判断充放电状态正确，则再加 1 分。

③回答错误为 0 分。

第 5 题，电解稀硫酸：有一未知浓度的稀硫酸，现用惰性电极做电解实验，随着时间的推移（时间无限长），问该电路的电阻如何改变？电流如何改变？如何检验，尝试画出电路电阻和电流随着时间的改变的变化图。

学生需要联系到化学的离子浓度和导电性，物理的电流与电阻以及欧姆定律，首先要利用电解池的知识判断离子的放电顺序，可以得到电解稀硫酸本质就是电解水，从而导致溶液浓度变大，离子浓度变大，溶液导电性增强，电阻变小，电流变大，但是长时间电解后，稀硫酸会变成浓硫酸，导致硫酸再溶液中大多数以分子形式存在，离子浓度减小，电阻变大，电流变小。学生不仅要分析前段反应原理，还要能敏锐地观察到浓硫酸的性质。该题没有涉及计算，只是定性分析，但是学生需要考虑到两个过程，这道题将离子浓度、导电性、电流和电阻进行联系整合，所以此题考察的能力水平为知识的整合与迁移水平 2。

①如果学生只回答电阻变小，电流变大，为 1 分。同时画出图，则再加 1 分。因为学生只看到了离子浓度变大却没想到浓硫酸的存在形式。

②如果学生回答电阻先变小后变大，电流先变大后变小，为 1 分。同时画出图，则再加 1 分。因为学生可以离子浓度、导电性和浓硫酸的性质结合起来进行思考。

③回答错误为 0 分。

(3) 概念的融合和量化能力水平 1 和 2 的试题分析：

第 8 题的第 2 小问：饱和氢氧化钠的密度为多少（单位 g/mL）？

这一问需要学生融合化学的溶液质量分数与物理的密度的概念，其转换公式为： $c = 1000\rho\omega / M$ ，通过这个公式可以直接计算出来，对溶质质量分数、溶液物质的量浓度与溶液密度进行量化，所以此题考察的能力水平为概念的融合和量化水平 1。

①如果能写出公式，则为 1 分。

②如果学生计算正确，则再加 1 分。

③写其他得为 0 分。

第 12 题的第 2 小问：求起跑过程中运动员能得到的加速度。（此过程为匀加速直线运动）

这一问需要学生化学的方程式的计算和物理的能量守恒定律和运动学规律等知识点，学生需要先通过方程式计算出 1.1g 葡萄糖氧化所提供的能量，再用所得出来的能量结合能量守恒  $E = (1/2)mv^2$  计算出末速度，再利用运动学规律  $v^2 = 2ax$  计算出加速度，这题需要熟练运用多个公式的计算，对能量、速度与加速度进行量化，所以此题考察的能力水平为概念的融合和量化水平 2。

- ①如果算出葡萄糖氧化提供的能量，则为 1 分。
- ②如果学生利用能量守恒计算出末速度，则再加 1 分。
- ③如果学生最终算出来加速度，则再加 1 分。
- ④回答错误为 0 分。

(4) 问题的解决能力水平 1 和 2 的试题分析：

第 10 题的第 6 小问：如果想要电解水的速率更快，可以采取什么措施？

这一问需要学生通过化学反应速率和电路知识认识，解决增大速率的问题。学生可以通过物理方法滑动变阻器减小电阻，从而增大电流，加快反应速率，也可以将电解的两个试管靠近，从而使离子移动的距离变短，相当于减小电阻长度，减小电阻，从而增大电流，加快反应速率。也可以通过加入其他的电解质使离子浓度增大从而加快反应速率，也可以提出其他合理方法，从而解决想增大电解速率的问题，所以此题考察的能力水平为问题的解决水平 1。

- ①学生回答出一种方法，则为 1 分。
- ②学生回答出两种方法，则再加 1 分。（合理答案均可）
- ③回答错误为 0 分。

第 13 题的第 3 小问：现在需要用电镀的方法在半径为 R 的铜球壳表面均匀镀上很薄的银层，在电镀槽中铜球是阳极还是阴极？另一电极是什么材料？若电流为 I，通电时间为 t，银的相对原子质量为 A，金属银的密度为  $\rho$ ，求镀层的厚度 d。（用本题中的符号表示）

这题需要学生通过电解池和电路相关知识，解决在铜球上镀银的问题。学生先需要通过电解池的相关知识得到：电镀时，通常以所镀的金属或合金作阳极，被镀制品作阴极。以含镀层金属离子的溶液作电解液因此要在铜球壳上镀银，应把铜球壳作为阴极，银作为阳极材料。再通过联系电子、电量和时间的公式  $Q = It$  得到第一个方程，再通过电子转移、体积、密度和物质的量计算得到第二个方程，最后进行联立得到最终答案。这题具有较高的综合性，利用到了知识点之间的联系、概念的相关计算，最后进行问题的解决，符合问题解决能力的综合内涵，虽然包含了其他几种能力，但是最终都是为了问题的解决，所以此题考察的能力水平为要素四水平 2。

- ①如果学生得出阴阳极，则为 1 分。
- ②如果学生列出第一个方程： $Q=It=Fn$ ，则再加 1 分。
- ③如果学生列出第一个方程， $\rho 4\pi R^2 d=nA$  则再加 1 分。
- ④如果学生最后计算出  $d$ ： $d=AI t / (4\pi R^2 \rho F)$ ，则再加 1 分。
- ⑤回答错误为 0 分。

### 4.3 测评工具的试测和优化

#### 4.3.1 样本选取和数据收集

在完成初步的诊断工具开发后，让导师进行检验，并且请一名中学物理教师和中学化学教师对测试卷的内容效度进行严格的审查。在此过程中，将特别关注测试卷中情境任务与项目问题的表述是否清晰明了，以确保所涉及的化学和物理知识难度适中且分布合理。同时，会审查预先设定的评分标准是否具备公正性和准确性，并评估每个项目所考察的能力要素是否符合预期水平。通过这一严谨的审查流程后，请两名化学课程与教学论的研究生同学进行测试，同时在中学让 55 名学生进行测试，回收有效试卷 50 份，测试卷回收率 91%，最后根据同学与学生的回答情况进行修改。

#### 4.3.2 各子能力数据收集

##### (1) 测评工具整体质量分析

##### ①整体质量统计

表 4-4 整体质量分析

	Measure	Error	Infit		Outfit		Seperation	Reliability
			MNSQ	ZSTD	MNSQ	ZSTD		
Persons	0.78	0.37	0.99	-0.10	1.01	0.10	3.23	0.91
Items	0.00	0.32	1.00	-0.10	1.01	0.10	3.84	0.94

在使用 Rasch 分析法时，认为当某项测验的难度水平被设置为中性（即零分）时，如果测试者的能力刚好与这个中性难度相当，那么便能够表明测试者与该测验项目间的匹配性是相对较高的。这个模型的分析结果不仅关注个体（即测试者），还包括了测验项目两个方面的数据。在这种情境下，“Measure”数值扮演了一个核心角色，它是以对数几率单元（logit）来量化的平均估算，用以衡量测试者的能力和测试项目的难度。对于 Rasch 模型而言，将项目难度默认设为零，进而使得测试者的 Measure 分数直接代表其能力水平的一种估计。若某个测试者的 Measure 分是正数，这可能意味着他/她具备较强的能力，抑或是测试项目偏易；反之，若是负数，则指向测试项目可能相对困难，或测试者的能力比较低<sup>[10]</sup>。通过对特定数据集的分析，以表格 4-4 为例，可以观察到测试者的 Measure 分数为 0.78，大于零，这表明该测试者在总体上拥有较高的能力水平。此外，从量表中观察到测试者与测验项目的 MNSQ（均方值）接近理想数值 1，而 ZSTD

（标准化残差）也接近理想数值 0，这些结果均标示出数据和模型之间拥有较好的贴合度。

在传统的测评理论中，分离度和信度这两个概念具有重要作用，它们主要用来判断测试对象在评估过程中能否保持其相对位置的稳定性。这不仅关乎个体的评价，还涉及到评估工具在区别不同水平难度任务的能力。一般而言，如果一个被测对象的分离度超过 2，并且其信度值趋近于 1，那么就说明这个对象和相关的任务都达到了较高的可靠性标准<sup>[49]</sup>。在进行的这项评估中，我们发现参与者的分离度为 3.23，显示出较强的区分度；而评估任务的分离度则为 3.84，这表明该评估方法能有效识别出不同能力层级的学生。与此同时，参与者的信度为 0.91，任务信度为 0.94，都非常接近理想的 1，反映出了这一评估体系的高度可信赖性。

项目与被试的模型误差（Error），为评估测量工具的精确度和可靠性提供了重要依据。模型误差作为衡量观测值与测量模型之间一致性的关键指标，其准确性直接关系到测量工具在 Rasch 模型检验中的表现。在本次测试中，模型误差分别为 0.37 和 0.32，处于可接受范围，这一结果验证了测量工具的高精确度和可靠性，同时也说明所收集的数据能够较好地真实反映学生的能力水平。

## ②项目拟合与误差统计

表 4-5 各项目拟合和误差分析

Items	Measure	Model S.E	Infit		Outfit		PT-Measure CORR.
			MNSQ	ZSTD	MNSQ	ZSTD	
13 (4)	2.47	0.26	0.80	-1.00	0.82	-0.90	0.65
10 (3)	2.26	0.25	0.76	-1.30	0.76	-1.30	0.73
11	2.22	0.30	1.08	0.50	1.07	0.40	0.43
10 (1)	1.58	0.23	0.92	-0.40	0.91	-0.40	0.66
13 (3)	1.47	0.31	1.12	0.70	1.16	0.80	0.40
3	1.11	0.26	0.77	-1.10	0.71	-1.20	0.62
14 (2)	1.05	0.23	1.09	0.50	1.09	0.50	0.59
9 (2)	1.00	0.28	0.93	-0.30	0.77	-0.30	0.59
13 (2)	1.00	0.28	0.77	-1.30	0.77	-1.30	0.69
10 (2)	0.97	0.22	0.60	-2.5	0.61	-2.40	0.83
4	0.91	0.22	1.65	3.00	1.66	2.90	0.31
5	0.87	0.18	0.22	-0.60	0.90	-0.50	0.72
9 (3)	0.84	0.25	0.87	-0.70	0.87	-0.60	0.66
8 (4)	0.79	0.29	1.04	0.20	1.05	0.30	0.51

7 (4)	0.77	0.23	0.93	-0.30	0.94	-0.30	0.64
10 (4)	0.52	0.32	0.74	-2.20	0.66	-1.60	0.66
12 (2)	0.25	0.20	0.98	0.00	0.97	-0.10	0.66
8 (2)	-0.01	0.23	1.02	0.30	1.24	1.00	0.54
6 (2)	-0.35	0.34	0.89	-0.70	0.80	-0.50	0.50
2	-0.46	0.26	1.33	1.70	2.68	4.7	0.30
1	-0.57	0.25	1.25	1.20	1.57	1.80	0.37
7 (1)	-0.64	0.25	0.97	-0.10	1.93	2.30	0.49
6 (1)	-0.73	0.36	1.00	0.10	0.78	-0.40	0.42
7 (3)	-0.81	0.29	0.92	-0.40	0.87	-0.50	0.57
7 (2)	-0.93	0.29	1.23	1.20	1.23	1.00	0.36
9 (1)	-1.08	0.28	0.96	-0.10	1.00	0.10	0.48
8 (5)	-1.71	0.46	1.14	0.60	0.81	0.00	0.24
12 (3)	-1.71	0.46	0.97	0.00	0.70	-0.20	0.34
13 (1)	-1.71	0.46	1.13	0.50	1.25	-0.60	0.19
8 (1)	-1.94	0.50	1.17	0.60	2.29	1.40	0.06
14 (1)	-1.94	0.50	1.06	0.30	1.06	0.40	0.22
12 (1)	-2.54	0.62	0.88	-0.10	0.39	-0.50	0.34
8 (3)	-2.99	0.74	1.09	0.30	1.15	1.10	0.04

通过对表 4-5 的详尽审查，可以得出项目的难度评分 (Measure) 覆盖了从-2.99 到 2.47 的 logit 值区间。在所研究的项目中，标号为 13 (4) 的题目表现出了最大的难度，而标号为 8 (3) 的题目则显示出相对最低的难度，由此可见项目间存在明显难度差异。综观表内数据，除了特定几个项目 (项目 4、项目 2、项目 7 (1)、项目 1 与项目 8 (1)) 的难度值超过 1.5 以外，大多数项目的拟合度 MNSQ 数值都稳定在 0.5 到 1.5 之间，这一区间被视作拟合度的合理标准。需要特别指出的是，项目的 Outfit MNSQ 指标能够敏感地反映出异常值。这暗示着所提及的项目中可能存在由于能力较低的考生意外答对或能力较高的考生错误作答的异常情况。关于 ZSTD 值，大多数项目的指标都集中在-2 和 2 之间的范围内，表明只要 MNSQ 值在接受的范围，ZSTD 的数值对于拟合度的评价影响并不显著。另外，考量到 PT-Measure CORR. (点测量相关系数)，其在所有项目中的正值反映出每个问题与整体测试之间保持着正相关性，提供了项目之间区分度良好的证据。总体分析结果表明了被试答题反应与预设模型预测的高度吻合，验证了模型的有效性。

### ③单维性检验

Rasch 分析方法它通过评估测验项目是否对应于单一心理特质尺度来检验测量工具的一维性。在心理学领域，通常假定个体拥有特定的隐性属性，而这些属性可以通过特定的测试项目来评估。当一个测验工具的项目都指向一个共同的潜在特质时，这个工具就被视为具有良好的一维性。一维性的确立对于确保一个测量工具的有效性至关重要，因为它表明该工具真实地反映了个体在某个特定领域的的能力。在本研究框架下，实现一维性的标准表明测试能准确地度量了个体的能力<sup>[61]</sup>。

如图 4-1 所示，Winsteps 软件输出的标准残差图展示了横坐标为项目难度，纵坐标为隐性因素的分布情况。在各字母代表的不同测试项目中，大部分项目落在横坐标[-0.4, 0.4]的区间内，符合单维性特征。值得注意的是，B-题目 9（1），C-题目 10（1）D-题目 10（3）的坐标未超出[-0.5, 0.5]的范围，仍可接受。然而，其余题目与这区间的距离较远，可能测量了其他能力，因此在后期的测试设计中，对这两个题目适当进行修改。

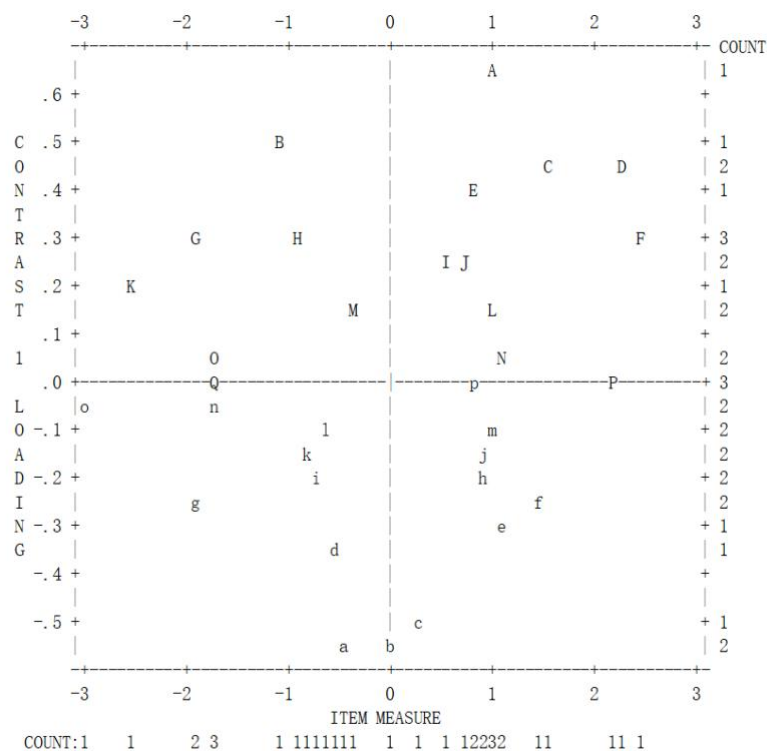


图 4-1 标准残差图

#### ④项目-被试对应

怀特图是一种基于 Rasch 模型的分析工具，其主要功能是将原始分数转换为 logit 分数。这种转换能够在统一的尺度上对项目难度和应试者能力进行比较和评估。进而可以利用这一工具对测试结果进行深入分析，从而对学生的跨学科能力进行精准评估。此外，通过怀特图，项目难度与应试者能力之间的关系得以直观地呈现<sup>[60]</sup>。

图 4-2 示意图展现了学生成绩的分布情况，其中每位学生用“X”标记。图中的中央线作为参考，显示了分数从原始值转换到 logit 值的比例尺。该线上标记着几个关键点：M 表示平均得分，S 和 T 则分别表示比平均得分高出 1 个和 2 个 logit 单位的得分水平。在所示的怀特图布局中，最左边的列代表难度因子与学生能力的对比数值。这一数值随着图表自底向顶部升高而增大，意味着难度和能力的提升。

在这种表示法中，项目的难易程度和学生的能力水平通过图表中的垂直位置展示。较为简单的任务和能力稍弱的学生被放在图表的下方，相反，难度较大的任务和能力较强的学生则被置于较高的位置。在这样的布局下，当一个学生的能力与某个任务的难度水平相当时，该学生有一半的几率能够正确解答该题目。

经过深入分析，本次测试中的 33 个项目在难度分布上展现出了良好的均衡性，其涵盖的范围亦相当广泛，足以适应不同能力层级的受试者。进一步观察受试者的能力分布，发现其呈现出一种中间多、两端少的态势，这与典型的正态分布模式相契合。此外，左侧受试者的分布与右侧项目的分布基本保持平衡，这充分证明了本次测试的工具项目能够精准地反映出学生在化学与物理这两个学科间跨学科能力的不同层级。

然而，尽管大部分情况都得到了很好的覆盖，但仍然存在一些空白区域。具体来说，左上角的高能力被试者并没有对应到相应的项目难度，这说明此次测试对于一些能力较强的学生来说可能过于简单。

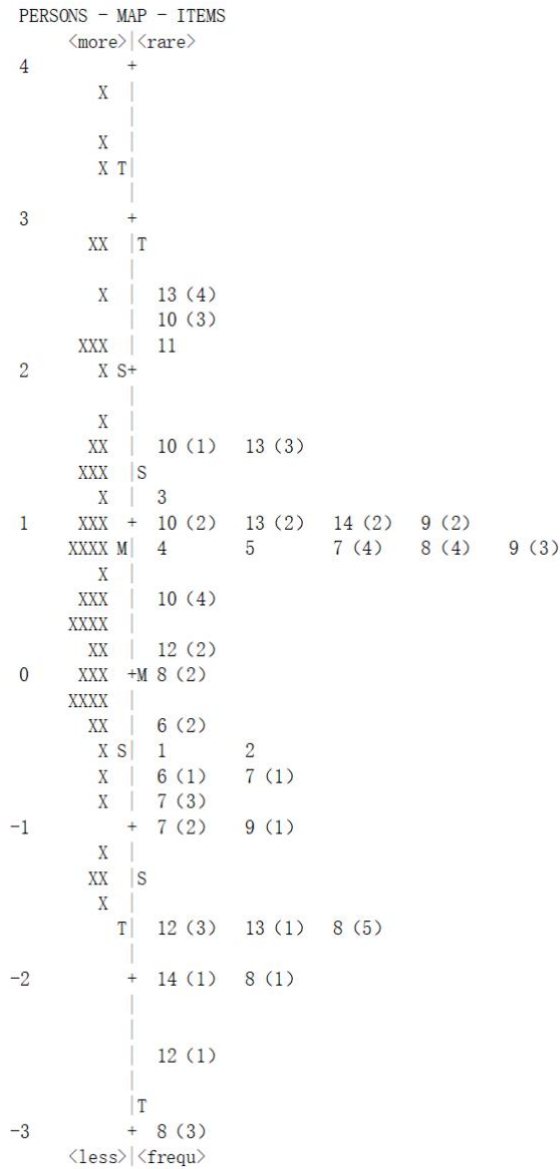


图 4-2 怀特图

(2) 信息的提取与转译能力指标分析

①整体质量分析

表 4-6 整体质量分析

	Measure	Error	Infit		Outfit		Seperation	Reliability
			MNSQ	ZSTD	MNSQ	ZSTD		
Persons	2.21	0.59	0.92	-0.20	1.05	-0.10	0.37	0.22
Items	0.00	0.53	1.00	-0.10	0.96	0.20	2.95	0.90

经过分析表 4-6 所展示的统计数据后，我们可以观察到参与者所获得的能力评分是 2.21，一个正数值，这显示出参与者具有相对较高的能力水平。此外，与评测项目相关的 Infit 和 Outfit 指数，包括 MNSQ 和 ZSTD，都非常接近于理想情况下的数值，这强调

了评测模型与实际测试结果之间存在良好的一致性和拟合度。尽管测试模型的拟合度较好,但从参与者的角度来看,存在一些值得关注的问题。首先,参与者的区分度只有 0.37,明显低于标准的 2 这一临界值;同时,可信度仅为 0.22,远未达到通常认为合格的 0.8 标准。这两个指标都指向了一个现象,即参与者的区分度与可信度指标需要被提高。在对项目质量的评估中,情况则显得相对乐观。项目区分度的数值是 2.95,超过了临界值 2,显示出能够较好地区分不同能力水平的被试者。项目的可信度为 0.90,也超过了接受的标准值 0.8,这再次暗示评测项目在区分度和可信度方面表现出色。总体来说,项目指标方面很理想,但被试者相关的指标则需要进一步改进和提升。

### ②项目拟合与误差分析

表 4-7 信息的提取与转译能力各项目拟合与误差分析

Items	Measure	Model S.E	Infit		Outfit		PT-Measure CORR.
			MNSQ	ZSTD	MNSQ	ZSTD	
13 (2)	3.85	0.37	0.77	-1.20	0.75	-0.80	0.82
7 (1)	0.73	0.28	0.69	-1.60	0.54	-1.30	0.67
13 (1)	-0.37	0.48	1.20	0.80	1.16	1.40	0.26
8 (1)	-0.62	0.51	1.22	0.80	1.48	1.80	0.16
14 (1)	-0.62	0.51	1.14	0.50	0.95	0.40	0.27
12 (1)	-1.25	0.63	0.83	-0.20	0.49	-0.10	0.30
8 (3)	-1.72	0.75	1.14	0.40	1.29	0.60	0.13

在对表格 4-5 所表示的数据进行分析后,我们可以观察到项目的难度等级是有明显区别的。具体来说,难度值浮动在-1.72 至 3.85 个 logit 单位间。这个发现揭示了各项测试在难度上的不同。同时,项目的 MNSQ 都保持在 0.5 到 1.5 之间,这个区间被认为是合适的,说明测试结果具有较高的可靠性。另外,ZSTD 指标的值也保持在了-2 到 2 之间,这进一步彰显了所得数据的稳定性。更进一步分析,点测量相关系数,即 PT-Measure CORR,全部呈现为正数,从这一点可以看出不同项目间存在着正向的关联性。这种关联性表明数据之间有着较为紧密的联系。基于以上各项指标分析和解读,我们可以得出一个结论:测试对象的反应与模型的预估结果在较高程度上是吻合的,显示出良好的一致性。这为该模型或测量工具的精确度提供了验证。

### ③单维性检验

根据 Rasch 模型的核心原则,测试的设计必须遵循单一维度原理,即仅通过评估学生的信息的提取与转译能力来解释测试结果。这意味着只有这一因素会影响学生的答题表现。通过分析图 4-3 的数据,我们发现多数测试题目的分布在[-0.4,0.4]的范围内,这与单一维度的要求是相符合的。然而,有两个具体的题目,分别是 7-题目 13 (2) 和 1-

题目 8 (1)，它们的数据表现并不符合这一标准，因此在后期的测试设计中，对这两个题目适当进行修改。

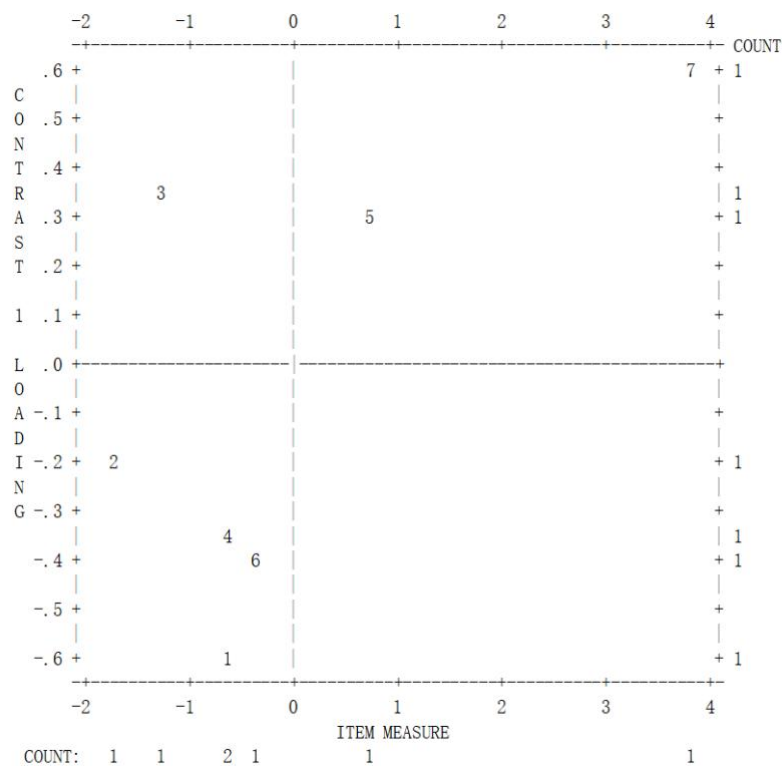


图 4-3 标准残差图

#### ④项目-被试对应

在图 4-4 的示意图中，一对“#”符号代表了两名学生，而每个“.”则代表一名学生。通过分析发现，大多数学生的能力水平超过了所给题目的平均难度。具体来说，顶端存在 6 名学生未能找到与之相匹配的题目，这表明了对于这组学生而言，题目的整体难度偏低。进一步观察学生能力的分布，我们可以看到其遵循一个较为典型的正态分布模式，绝大部分学生的能力集中在中等范围，而能力较强或较弱的学生数量则明显较少。

在具体项目中，题目 13 (2) 位于分布图的最上端，表明该题的难度最大；而题目 12 (1) 则位于最下方，说明该题的难度最小。整体来看，试题的难度分布与预设水平基本相符。然而，值得注意的是，预设 7 (1) 的难度水平本应低于 13 (1)，但实际情况并非如此。因此，后续将考虑对这两道试题进行相应的调整。

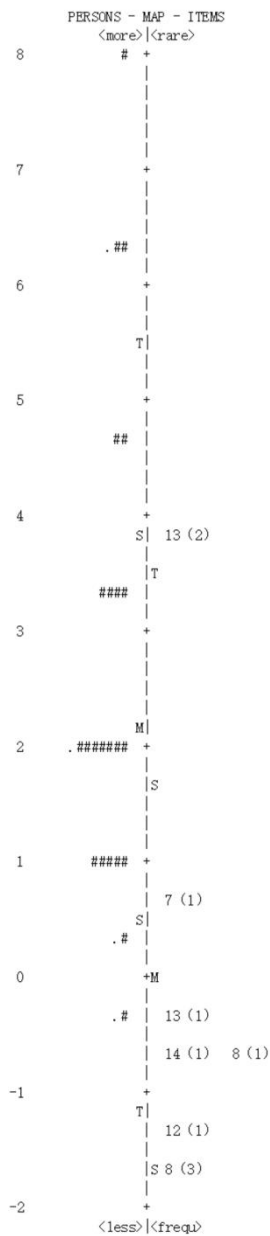


图 4-4 怀特图

(3) 知识的整合与联系能力指标分析

①整体质量分析

表 4-8 整体质量分析

	Measure	Error	Infit		Outfit		Seperation	Reliability
			MNSQ	ZSTD	MNSQ	ZSTD		
Persons	0.80	0.45	1.01	0.00	0.99	0.00	1.37	0.65
Items	0.00	0.28	1.00	0.20	0.98	-0.10	2.51	0.86

经过分析表 4-8 中所提供的数据，我们可以发现受试者的能力估计值达到了 0.80，这一正值指示了受试者具有较高的能力水平。进一步观察受试者与测试题目之间的相互

作用，通过 Infit 和 Outfit 统计指标的 MNSQ 以及 ZSTD 接近理想标准，从而可以确认所用评估模型在拟合实际测试数据方面的有效性。不过，也有一些值得注意的地方。受试者的区分度只有 1.37，未达到 2 的基准线，同时其信度也仅为 0.65，未能达到 0.8 的理想阈值。这两个指标显示出受试者之间的区分度和信度只能说是一般。相较而言，测试题目本身的表现要好得多。题目的区分度为 2.51，超过了 2 的临界值，而信度更是达到了 0.86，超出了 0.8 的优良水平。这表明我们所使用的题目集在区分受试者和保持评估一致性方面是有效的。

### ②项目拟合与误差分析

表 4-9 知识的整合与联系能力各项目拟合与误差分析

Items	Measure	Model S.E	Infit		Outfit		PT-Measure CORR.
			MNSQ	ZSTD	MNSQ	ZSTD	
5	1.08	0.26	0.77	-1.10	0.72	-1.10	0.65
3	0.91	0.22	1.55	2.50	1.53	2.40	0.32
4	0.87	0.18	0.68	1.80	0.71	-1.60	0.80
8 (4)	0.87	0.29	0.89	-0.50	0.87	-0.60	0.58
6 (2)	-0.28	0.34	0.90	-0.60	0.82	-0.60	0.51
2	-0.28	0.25	1.14	0.80	1.32	1.40	0.45
1	-0.49	0.25	1.04	0.30	1.01	0.10	0.50
6 (1)	-0.65	0.36	0.78	-1.30	0.58	-1.30	0.59
7 (2)	-0.86	0.29	1.10	0.60	1.09	-0.50	0.43
9 (1)	-1.00	0.28	1.12	0.60	1.21	0.70	0.41

经过对表 4-9 中的数据分析，可以观察到项目难度指标呈现在-1.00 到 1.08 这个数值区间内，说明不同的项目难度并非均质化，彼此之间是有区别的。对于大部分项目来说，他们的 MNSQ 值落在 0.5 到 1.5 这个认为是合适的区间内，意味着这些项目在测量上表现出了较为一致的性质。遗憾的是，题目 3 的 MNSQ 值出现在合理区间之外。此外，除了题目 3，大多数的 ZSTD 适中地分布在-2 到 2 之间，证明了数据在一定程度上稳定性。所有项目的点测量相关系数 PT-Measure CORR 均为正，显示了项目间健康的相互关联和数据之间的内在一致性。总的来说，这些结果表明，在大多数情况下，应试者给出的答案与模型预测之间保持较高程度的一致性。

### ③单维性检验

在应用 Rasch 模型进行数据分析时，该模型的基本预设是单一能力维度的测量。这意味着学生如何回答问题应主要受他们知识的整合与联系能力的影响。依照图表 4-5 的数据分布，我们观察到大多数测试条目的得分集中在[-0.4, 0.4]的范围内，这表明测试





Persons	-0.26	0.65	0.97	-0.10	1.00	-0.10	2.29	0.84
Items	0.00	0.20	1.00	0.10	1.00	0.00	3.61	0.93

通过分析表 4-10 中提供的数据，我们可以发现参与者的能力得分是-0.26，这个负数的结果指出相对于测试的标准，参与者的能力层次偏低。进一步观察到，参与者和测试题目的 Infit 和 Outfit 统计指标数值都非常接近理论上的完美指数，这表示了我们的模型与实际情况之间有很好的吻合性。此外，参与者的区分力指数为 2.29，超过了 2 的阈值，同时可靠性系数达到了 0.84，超出了 0.8 的可接受标准，反映了参与者表现的一致性与可靠性都是相当可观的。而且，针对测试题目而言，其区分力指数达到 3.61，显示出比参与者更高的可辨识能力，并且其可靠性系数也达到了 0.93 的高水准，这进一步证实了测试题目在区别不同能力水平的参与者方面表现出的优秀性能和高度可信性。

### ②项目拟合与误差分析

表 4-11 概念的融合和量化能力各项目拟合与误差分析

Items	Measure	Model S.E	Infit		Outfit		PT-Measure CORR.
			MNSQ	ZSTD	MNSQ	ZSTD	
10 (3)	1.49	0.27	0.84	-0.80	0.81	-0.90	0.73
11	1.39	0.32	1.16	0.90	1.33	1.40	0.46
10 (1)	0.68	0.25	0.95	-0.20	0.93	-0.30	0.70
13 (3)	0.53	0.33	1.23	1.10	1.24	1.00	0.47
14 (2)	0.06	0.25	1.10	0.60	1.08	0.50	0.66
9 (2)	0.00	0.30	0.94	-0.20	0.97	-0.10	0.65
10 (2)	-0.03	0.23	0.74	-1.40	0.72	-1.60	0.80
12 (2)	-0.89	0.22	0.99	0.00	0.92	-0.30	0.73
8 (2)	-1.14	0.25	1.10	0.60	1.07	0.30	0.62
7 (3)	-2.10	0.31	1.00	0.00	0.92	-0.20	0.58

参考表 4-11 的统计信息，我们观察到测试项目的难度水平分散在-2.10 至-1.49 logit 之间，这表示不同测试项目间存在难度上的差异。同时，项目的 MNSQ 值都落在 0.5 到 1.5 的接受区间，指示出测评结果具备高度的一致性。绝大部分的 ZSTD 值位于-2 到 2 之间，这一范围被视为数据稳定性的标准，从而印证了数据的可靠性。此外，点测量相关系数 PT-Measure CORR 均呈现出正数，反映了测试题目间有较强的相互关联，表明数据间的内部联系紧密。基于这些观察，我们可推断，参与者的响应与研究模型所做的预期结果之间展现出了高度的匹配度。

### ③单维性检验

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/178116113005006143>