第 31 卷第 4 期
2022 年 8 月
数 学 教 育 学 报
JOURNAL OF MATHEMATICS EDUCATION
Vol.31, No.4
Aug., 2022

双维多水平数学建模能力测评框架的构建

祖　丹[1]，丁　锐[1]，孔凡哲[2]

（1．东北师范大学 教育学部，吉林 长春　130024；2．中南民族大学 教育学院，湖北 武汉　430074）

摘要：数学建模能力是数学能力的核心要素，数学建模能力的形成对学生数学关键能力的培养和提升都起到了重要作用．基于数学建模的过程性特征，从纵横两个角度，构建了双维多水平的数学建模能力测评框架．双维分别是覆盖广度和覆盖深度，其中覆盖广度是指建模过程的完成情况，覆盖深度是指模型假设能力、模型构建能力以及模型检验能力的水平．基于 SOLO 分类理论和“关系—表征复杂性模型”，构建了覆盖深度下各建模子能力的评价标准．专家咨询和测验结果均证明了双维多水平数学建模能力测评框架及评价标准具有较高的信效度，能够较好地评价学生建模能力水平．

关键词：数学建模；数学建模能力；测评框架；覆盖广度；覆盖深度

中图分类号：G40–034　文献标识码：A　文章编号：1004–9894（2022）04–0056–06

引用格式：祖丹，丁锐，孔凡哲．双维多水平数学建模能力测评框架的构建[J]．数学教育学报，2022，31（4）：56–61．

1　问题提出

《普通高中数学课程标准（2017 年版）》中强调数学教育不仅要帮助学生掌握现代生活和进一步学习所必需的数学知识、技能、思想和方法，还应能引导学生学会用数学眼光观察世界，会用数学思维思考世界，会用数学语言表达世界[1]．数学建模是一种认识世界的重要工具，立足于运用数学语言去描述现实世界中事物的本质关系和规律，学生对数学建模的理解和掌握程度影响其解决现实问题的思维和行为方式．目前，利用评价来全面了解和诊断学生数学建模能力的发展状况，进而提供针对性的数学支持，是培养学生数学建模能力的一种较为可行的思路．在此之前，如何构建科学、合理的测评框架是现阶段亟待解决的问题．

以往的研究中，大部分研究者致力于开发数学建模能力的评价标准，对数学建模的总体能力进行评价．如 Herbert 等人将数学建模分为 3 个水平：识别和理解建模（recognition and understanding of modeling）、独立建模（independent modeling）和建模元反思（meta-reflection on modeling）[2]．朱娅梅将数学建模能力分为再现、联系和反思 3 个水平[3]．少有的关注数学建模过程性子能力的研究，如 Zöttl 也仅是关注了子能力是否被“激活”[4]，并不关心学生的子能力水平．基于此，在探讨数学建模能力构成要素的基础上，构建了双维多水平数学建模能力测评框架，该框架不仅关注了学生数学建模过程的执行情况，同时还能测量各建模子能力的水平，最后研究采用定量的方法验证了该框架的合理性和可操作性．

2　研究设计

首先，采用专家咨询法验证测评框架及评价标准的信效度．研究者选取了 4 位专家（两位数学教育专家，两位大学数学建模课程教师）作为咨询专家，了解他们对测评框架及评价标准的认同度．

其次，以评价标准为依据，计算评分者的一致性．基于分析框架，结合 PISA 测试和《义务教育阶段数学课程标准（2011 年版）》，自编数学建模能力测试卷，并进行测试．选取 H 市五~八年级学生作为被试，共发放 236 份问卷，回收 236 份，其中有效问卷 213 份，有效率 90.25%．随机选取 40 份试卷，要求 3 位数学教育研究者分别独立对 40 份试卷评分，并计算评分者一致性（肯德尔系数）．

最后，针对测验结果，利用方差分析计算不同年级的学生在各维度上的得分差异，以此验证测评框架的可操作性和合理性．

3　数学建模能力的内涵及测评框架的构建

3.1　数学建模能力的内涵和要素构成

对数学建模能力评价的首要任务是明晰数学建模能力的内涵及构成要素．通过对已有研究的梳理，国际上对数学建模能力构成要素的划分大体可以分为以下两个视角：一种是从纵向的角度，对数学建模过程进行分解；一种是横向的角度，分析不同的数学建模阶段所需的所有子能力．

已有对数学建模过程的划分，影响最广的是 Blum 等人的研究．早期 Blum[5]认为数学建模是从现实情境到现实模型、数学模型、数学结果最终回到现实情境的一个循环过程（如图 1），后来考虑到现实情境的复杂性，他们在原有框架中加入“情景模型”和“现实结果”，用以分析个体理解任务的能力，将 4 个阶段细化成 6 个阶段．之后的研究者选取不同的视角，对 Blum 的框架进行调整，如 Greefrath 增加计算机技术环节[6]，Borromeo 从心理学的视角描述数学建模周期[7]，Galbraith 将数学建模描述为一种双向循环的过程[8]等．PISA 认为积极参与问题解决或数学建模的学生应该经历：数学化、应用、阐释与评估 3 个步骤．简要说来，“数学化”指运用数学的方法表示现实问题．“应用”指利用数学概念、程序、事实和工具来推导数学解．“阐释与评估”指反思解决方案或结果，并在现实背景下解释它们[9]．

对应建模过程各阶段，Blum 和 Kaiser 等人将数学建模能力划分为：解决实际问题并构建现实模型的能力、构建数学模型的能力、求解数学模型能力、在现实模型或现实情境

收稿日期：2022–02–09
基金项目：教育部人文社会科学研究规划基金 2019 年度一般项目——小学生数学核心概念学习进阶的构建与诊断（19YJA880007）
作者简介：祖丹（1991—），女，吉林珲春人，博士生，主要从事数学教育研究．

中解释数学结果的能力以及验证结果合理性的能力[10]．

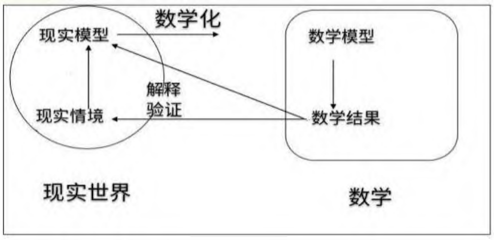


图 1　数学建模循环过程

除了依据数学建模进程划分的一系列子能力，研究者发现，学生在建模时，并不是完全遵照建模进程，有时候还会出现逆行的现象，如从数学模型返回到现实模型．从心理学的角度来看，他们认为这是学生的自我监督和反思．基于此，Vorhölter 等人认为元认知能力也是重要的建模能力[11]．除了认知和元认知能力，Biccard 认为学生在建模时还需要情感能力，这里情感能力主要指的是信念[12]．Maab 认为成功完成数学建模任务还需要有运用数学解决问题的意识、解题的方向感以及论证能力等[10]．随着研究的深入，除了个体的数学建模能力，Biccard 和 Wessels 还研究了团队建模能力[13]．尽管从横向角度分析数学建模能力的研究越来越多，但随着子能力的增加，建模能力的构成要素越复杂，也就越难以测量和评价．因此，有必要进一步分析建模能力的基本构成要素，并构建可供测评的建模能力框架．

基于 Blum 和 PISA 的研究，研究从纵向的角度将数学建模过程分解为模型假设、模型构建、模型检验．下面进一步解释这 3 个步骤．数学建模的核心是刻画事物之间的数量关系，而如何刻画事物之间的关系全凭人们的想象，可以将这个想象称之为模型假设[14]．模型假设是数学建模的第一步，指建模者对情境中的信息进行理解、编码，进而针对问题目标筛选主要变量，并对主要变量的关系做出合理化假设的过程．其次，模型构建是个体选择和运用数学概念、公式和定理近似地刻画现实中一类问题的本质和规律，并得到数学解的过程．最后，模型检验是使现有模型逐渐贴近真模最重要的一步，是对结果、模型和解决方案进行反思的过程，其最终目的是根据反馈信息评价和调整模型．对应这 3 个阶段，可以将数学建模能力划分成 3 个过程性子能力（如表 1）．

表 1　数学建模过程能力分解

数学建模子能力	具体描述
模型假设能力	分析和识别现实情境中的主要变量，并对它们的数量关系做出合理化假设的能力．
模型构建能力	选择和运用数学概念、公式和定理等构建数学模型来描述变量之间的关系，并得出数学解的能力．
模型检验能力	对数学建模过程中的结果、解决方法和数学模型等进行评价与反思的能力．

3.2　数学建模能力测评框架构建

Box 和 Draper 认为模型是对现实的一种简化抽象，所有的模型都是错误的，但有些是有用的[15]．事实上，建模者对数量关系的刻画有可能是“正确的”，也可能是“错误”的．这种“错误”不是一种“失败的解决方案”，而是对现实的片面解释．

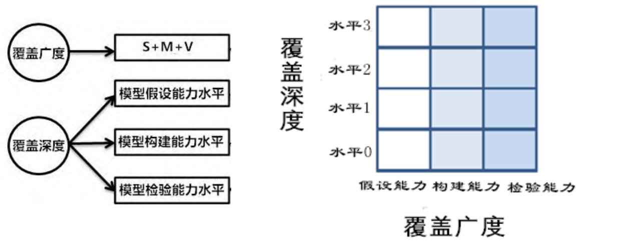
在一次测验中，研究者要求学生回答“牛吃草问题”：草场里有牛喜欢吃的草，23 头牛可在 30 天内吃完，33 头牛可在 20 天内吃完，问这些草可供 43 头牛吃几天？L 学生认为可以吃 10 天，并且给出了算式“牛的数量−23=30−天数”，研究者问：“你觉得你的答案正确吗？”L 学生回答：“我将 23、30 和 33、20 分别带进去，等式成立，所以我认为是正确的．”此时该学生完成了数学建模的 3 个步骤，但因为忽视每天牧草都会生长，且没有考虑每头牛吃草速率不同等现实意义，并没有得到“正确的”数学解．事实上，数学建模问题具有情境开放、方法和结果多样等特点，简单地根据数学结果正确与否来评价学生的数学建模表现并不恰当．

正如上述例子，虽然这个学生没有构建“正确”的数学模型，但相对那些完全没有回答的学生来说，该学生执行了数学建模的部分过程，因此，出于对这种现实情况的考虑，研究认为学生数学建模活动的完成度可以作为衡量数学建模能力水平的重要指标．进一步 Jesen 认为可以用覆盖度（degree of coverage）来描述学生的数学能力水平．个体能够自发激活数学能力的子程序的数量越多，则数学能力的覆盖越广，学生的能力水平越高[16]．从过程的角度来看，数学建模过程中各步骤之间存在递进关系，前一步是后一步的基础．根据 Jesen 的研究，结合数学建模的过程性特征，个体在一次数学建模活动中所能激活的过程性子程序的数量，即对应数学建模活动的完成度．因此，个体激活的子程序越多，则表示经历的数学建模步骤越多，将这种从水平的方向评价学生数学建模能力的维度称之为覆盖广度（如图 2（*a*）所示，其中，S 表示模型假设步骤，M 表示模型构建步骤，V 表示模型检验步骤）．

这种利用覆盖广度评价数学建模能力的做法尽管简单易行，但也存在一定的问题．如在研究“玉米的生长高度与时间的关系”问题时，建模者们分别用线性模型和非线性模型对数据进行拟合，虽然均能够得到数学解，但两个模型对现实的解释程度有高低之分．从结果的角度来看，学生数学建模得到的结果对现实的解释度越高，其数学建模能力越强．需要注意的是，个体在建模各阶段的作答表现均会影响最终的建模结果，而学生各子能力的水平决定了其在各阶段的作答表现．因此，可以通过描述各子能力的水平来判断个体的数学建模能力水平，将这种从垂直的方向评价学生数学建模能力的维度称之为覆盖深度．数学建模能力评价的覆盖深度维度涵盖 3 个二级指标：模型假设能力水平、模型构建能力水平、模型检验能力水平，如图 2（*a*）所示．

总之，可以从“过程的完成度”和“结果的有效性”两个方面描述学生的作答表现．对应于这两方面，研究将覆盖广度和覆盖深度作为衡量个体数学建模能力的标尺．

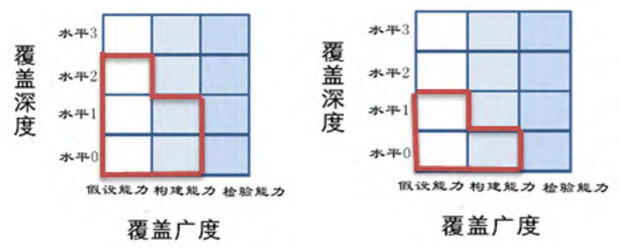
如图 2（*b*）所示，覆盖广度和覆盖深度分别从横向和纵向两个维度评价学生的数学建模能力．从横向上来看，每多覆盖一种颜色的格子，意味着个体能激活更多种类的子能力；从纵向上来看，每列格子数越多说明该列建模子能力水平越高．



（*a*）覆盖广度维度与覆盖深度维度 （*b*）数学建模能力测评框架

图 2 数学建模能力分析框架

图 3 为学生作答表现的两种示例，封闭图形区域表示学生在各维度上的水平．图 3（*a*）中，A 学生在各维度的表现分别为：完成了建模的两步，模型假设能力处于水平 2，模型构建能力处于水平 1．图 3（*b*）中，B 学生在各维度的表现分别为：完成了建模的两步，模型假设能力处于水平 1，模型构建能力处于水平 0．



（*a*）A 学生作答表现 （*b*）B 学生作答表现

图 3 建模能力测评框架应用实例

4 数学建模能力评价标准构建

根据上述建模能力的评价框架，下面分别说明如何从覆盖广度和覆盖深度两个维度评价学生的数学建模能力水平．

4.1 覆盖广度维度的评价标准

覆盖广度维度测量的是个体数学建模过程的完成度，这是从过程的角度对数学建模能力进行评价．具体来说，可以通过统计个体在一次数学建模活动中累积完成的步骤来评价学生在覆盖广度维度上的表现，即能够自觉完成数学建模过程的步骤越多，则覆盖广度维度得分越高．

4.2 覆盖深度维度的评价标准

覆盖深度维度是测量个体数学建模各子能力水平，这是从结果的角度对数学建模能力进行评价．基于辛自强的关系—表征复杂性模型和 SOLO 分类法，研究分别对数学建模各子能力水平进行划分，并使用例 1 来说明学生在不同子能力上的水平表现．

例 1 姚明身高 2.26 米．某次姚明受邀参加“一日学生”的活动，与同学们一起上课．请问姚明需要多高的课桌（桌面高）和座椅（椅面高）？

（1）要解决这个问题，你认为都需要考虑哪些量？

（2）写出你会用到数学式子并计算．

（3）你能在数学或现实背景下检验你的结果吗？

（4）现在有一个人的身高为 x（单位：cm），他需要的椅子面高为 y（单位：cm），你能写出身高 x 与椅面高 y 的关系式吗？

此题为开放式问题，结果并不唯一，允许学生自由寻找解决方案，选择建模方法．

（1）模型假设能力的水平划分．

现实情境往往是杂乱无章的，建模者在模型假设时，需要在大量的无关变量或弱相关变量中，识别关键变量并且对这些关键变量的关系进行数学化表征．例如，在研究“植物的生长期问题”时，日照时长、灌溉量、土壤等因素均会影响植物的生长期，这里建模者无需考虑全部的影响因素，而是根据研究目的控制一些无关变量，仅考虑主要因素之间的关系即可．个体具体表征了多少个变量且这些变量关系的复杂程度受限于建模者的模型假设能力水平．

辛自强提出的“关系—表征复杂性模型”从表征广度和表征深度两方面描述了个体表征的水平[17]．“关系—表征复杂性模型”是从问题本身的结构入手，判断学生的表征水平，此时问题的难度和考查的知识点是由出题者控制的．而在解决数学建模问题时，往往需要学生针对问题目标自行提出数学问题，此时题目内部各变量及其关系取决于建模者．因此，个体的模型假设能力水平可以通过其对现实问题的数学化表征的复杂程度来确定．

研究基于“关系—表征复杂性模型”，结合 Biggs 提出的 SOLO 分类理论，将模型假设能力划分为 4 个水平（表 2）．

表 2 模型假设能力的水平划分及学生表现

水平	各个水平的具体描述	学生表现
水平 0	思维较为混乱，只能识别个别变量，没有形成问题结构．	仅提及姚明的身高，姚明所需的椅子，并未表述两者之间的关系．
水平 1	能围绕问题目标的子目标识别部分变量，或围绕问题目标识别较少变量，形成较为简单的问题结构．	姚明的身高，身高与椅子比例，姚明所需椅子高度．
水平 2	能围绕问题目标识别较多变量，形成多个独立简单的问题结构．	姚明的身高，姚明所需椅子（桌子）高度，我的身高，我的椅子（桌子）高度．
水平 3	能围绕问题目标整合思考，识别全部或大部分变量，形成较为复杂的问题结构．	姚明的身高，姚明所需椅子（桌子）高度，平均的身高，平均的椅子（桌子）高度，桌子与椅子差比．

模型是对现实的简化，它描述了某一现象的本质特征，控制或忽略其它无关特征．一般来说，学生识别变量和表征它们的关系是同时进行的．如果学生的假设过于简单，意味着他们提取的关键变量较少，刻画的数学关系较为简单．

（2）模型构建能力水平划分．

将假设付诸实际的过程，需要大量使用数学中已经成熟的运算法则和推理法则．在对主要变量的关系进行结构化处理时，不同模型构建能力水平的建模者选择的建模思路和策略有所不同，这种差异最终会影响模型的有效性，因此，对个体模型构建能力的描述可以从“模型的有效性”入手．所谓“模型的有效性”包括：模型的适用范围和对现象的解释度．模型适用范围的有限性是数学模型的基本性质，这个适用范围通常表现于模型的假设前提，表现于模型的初始值，或者表现于对模型参数的某些限制[14]．作为沟通数学与现实的桥梁，除了数学价值，人们还关注模型的应用价值，适用性强的模型具备强迁移性，改编参数值结果相差不大，此时模型具备较高的应用价值．模型对现象的解释度指数学模

型对现象的近似刻画的程度，即模型与真模之间的距离，构建的模型越贴近于真模，其对当前现象的解释度就越好．研究以“模型的有效性”作为测量模型构建能力的指标，基于 SOLO 分类理论将个体模型构建能力划分为 4 个水平（表 3）．

表 3　模型构建能力的水平划分及学生表现

水平	各个水平的具体描述	学生表现
水平 0	提出较为混乱的建模思路，不能构建数学模型或者模型、结果不符合当前现象．	姚明与我的身高相差 38 cm，姚明需要的桌子高为：$\frac{2.26+0.38}{2}=1.32$（米）．
水平 1	构建的数学模型较为片面，结果不能很好地解释当前问题．	$\frac{2.26}{2}=1.13$（米）．
水平 2	构建的模型是能解释一种现象的特例模型，结果能较好地解释当前现象．	我的身高 165 cm，椅子座面高 40 cm．设姚明的椅子座面高 x cm $\frac{165}{40}=\frac{226}{x}$ 解得 x=54.8．
水平 3	结果不仅能很好地解释当前现象，还能较好的描述并解释一类现象，且构建的模型具有一定的适用范围．	利用分段函数描述身高与椅子面高度的关系．例如：全班身高范围为 158~172 cm，身高差值为 14 cm，平均身高为 165 cm．椅子面高度为 40 cm，全班同学使用适宜．保持这个增幅，身高范围在 173~187 cm 的同学，通过计算可知，椅子面高为 44 cm 适宜．以此类推，可以构建函数关系式：$y=4k+40$．当 $k=4$ 时，身高的区间为 $218<x<232$，姚明的身高正好落在这个范围，因此，姚明需要 56 cm 的椅子较适宜．

值得注意的是，构建模型时需要平衡模型的适用范围与解释度．有的模型具有非常广泛的适用范围，但是对现象的解释度较低；有的模型过于具体，虽然对当前现象有较好的解释，但不具有迁移属性，应用价值较低．因此，在考察学生的模型构建能力水平时，要综合考虑其构建模型的一般性与解释度．

（3）模型检验能力水平划分．

由于现实情境中信息的不确定性，数学建模的“目标”与“结果”并不是单一映射，而是保持一种“动态平衡”．数学建模是一个迭代的过程，在反复迭代过程中模型对现象的描述、解释和预测的准确性越来越高，建模结果也越来越贴近真值．因此，需要对数学建模进行检验，以便及时调整解决方案或模型，使其更好地解释现象．

PISA（2021）认为阐释与评估素养是指个体反思数学解决方案、结果或结论的能力．具体来说，可以从以下几个方面对数学结果进行解释、应用和评估：解释图形或图表信息；根据问题情境评估数学结果；回到现实世界解读数学结果；在现实问题背景下评估数学解决方案的合理性；理解现实对数学运算程序或建模结果的影响，并为调整和应用结果做出相关判断；在给定背景下解释数学结果或结论是否具有意义；理解数学概念和数学解的限定范围；评价和识别数学模型的适用范围；运用数学思维和计算做出预测，并比较和检验解决方案[9]．就数学建模而言，Blum 认为模型检验包括检验、反思、分析、评价模型和建模结果[18]．

根据 Blum 和 PISA（2021）的研究，就检验对象而言，可以分为对数学结果的检验和对解决方案或模型的检验．从检验的范围来看，可以在数学领域或现实背景下对模型和结果进行检验．研究中模型检验包括 3 部分内容：在数学背景下检验结果的正确性、在现实背景下检验结果的合理性、评价解决方案的合理性和模型的局限性．基于 SOLO 分类理论将个体模型检验能力划分为 4 个水平（如表 4）．

表 4　模型检验能力的水平划分及学生表现

水平	各个水平的具体描述	学生表现
水平 0	不能检验或能进行直觉检验，但不能解释原因．	我直觉感觉可能不太准确，但说不明白．
水平 1	采用重新算一遍或者带入法等方法检验数学计算结果的正确性．	重新验算了一遍，结果正确．
水平 2	能运用其它数学方法或在给定问题背景下解释、评估数学结果．	通过比例计算得到姚明的椅子高为 54.8 cm，椅子面占总身高的 24.2%，人坐下弯曲小腿，所以椅子长与小腿长差不多，人的小腿约占身体的 25%．可以看出得到的结果相差不多．
水平 3	在验证结果正确性的基础上，在现实情境中评价解决方案的合理性和模型的局限性，并提出改进方案．	我觉得我现在的结果太特殊了，我自己的身高并不能代表所有人的身高，全班同学的身高不同，但是用同一套桌椅，因此，可能需要分别计算一段身高范围的平均身高与对应桌椅高的比例．

事实上，在进行检验时，为使模型更贴近真模，建模者一般从结果出发，对得到的结果、解决方案以及模型提出合理的质疑，并以此作为修正模型的基础．

5　双维多水平数学建模能力测评框架及评价标准的验证

首先，专家咨询的结果显示，专家们对各个指标和标准的认同度较高，均达到了“认同”以上的程度（1 为非常不认同，5 为非常认同），说明该测评框架具有较好的效度（如表 5）．根据专家对各维度、指标的认同度，计算肯德尔系数为 0.787，说明专家们的意见较为一致，具有良好的信度．

表 5　专家认同度

专家咨询内容	认同度
一级维度	4.25
二级指标	4.75
评价标准	4

其次，评分者一致性的结果显示，3 位评分者对试卷评分的肯德尔系数均高于 0.7，说明不同评分者利用同一个评价标准的评分结果一致性较高，该评价标准具有较好信度．

最后，根据学生在此次测验中的表现，进一步分析该测评框架和评价标准的区分度以及学生的表现．方差分析结果显示，不同年级的学生在各维度上的表现具有显著差异，且在各个水平上均有学生分布，说明测评框架和评价标准能够分析不同年级的学生数学建模能力发展状况，具有较好的区

分度和合理性．

具体分析学生在各维度的表现，不同年级学生在覆盖广度维度的得分有显著差异（F=5.048，P=0.008，效应量=0.27）．具体来看，62%的学生有能力完成两步及以上的数学建模过程，这说明大部分中小学生能够围绕问题目标进行模型构建活动，初步具备了一定的建模意识．

但从覆盖深度维度来看，如图 4 所示，大部分学生在各建模阶段上的表现并不是很好．具体来说，不同年级的学生模型假设能力水平差异显著（F=4.633，P=0.004，效应量=0.26）．在面对情境开放的问题时，大部分学生能够基于生活经验对关键变量及其关系进行简单描述．

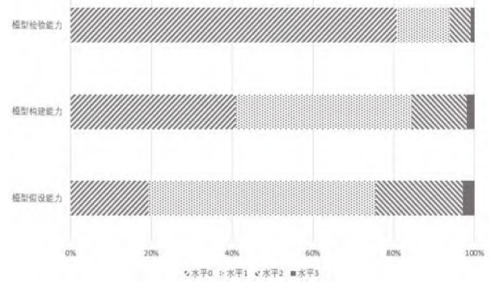


图 4　学生的各子能力水平统计结果

其次，不同年级的学生模型构建能力水平有显著差异（F=5.886，P=0.001，效应量=0.29）．总的来说，学生在解决开放问题时，在模型构建阶段上的表现较差，其中，59%的学生处于水平 1 及以上，且仅有 16%的学生处于水平 2 及以上，可见大部分学生仅采用直接估计等方法构建模型，鲜有学生使用列比例式、三角函数等方法构建模型．实际上，学生在日常学习中接触的大部分问题是已经“去现实化”的应用问题，因此，中小学生已经形成了解“应用题”的惯性思维，只要是老师提出的或者教科书上的问题就是可以解的且有意义的；每个题目都有单一的、精确的解，题目中给出的数字必须使用；不必在意相关内容是否违背常识[19]．而这种对某一知识、技能的操作性训练，本质上还是在数学领域内进行解题活动，并没有联系实际．因此，当遇到与现实情境紧密相连的开放问题，尤其是如例 1 这样仅给出“2.26”一个数字信息的问题时，大部分学生秉持可行性原则，尽可能仅利用现有数据，根据生活经验直接估计桌子的高度（如图 5），只有较少的学生能够想到利用自己的桌高、椅高等隐含信息求解．

最后，不同年级学生的模型检验能力水平有显著差异（F=13.840，P=0.000，效应量=0.44）．学生的模型检验能力水平较低，81%的学生还处于水平 0，即没有进行检验或仅进行了直觉检验．有 13%的学生处于水平 1，这部分学生更关注数学领域内的运算程序以及数学结果的正确性，并不关心计算的数学结果是否符合现实情况，比如：

主试：你检验了哪些内容?

被试：我就回去看了一下结果跟题目要求是不是一样(相符)．

主试：你都知道哪些检验方法?

被试：我们老师教我们做完题要再验算一遍，要不把结果带到题目中再看一遍数对不对．

可见，教师在日常教学中更强调检查计算结果的正确性，较少要求学生在现实背景下反思建模方案和结果的合理性，这可能是导致学生模型检验能力水平低的重要原因．近年来，数学课标对学生的数学建模能力提出了越来越高的要求，但对建模过程中的“检验”与“应用”步骤的重视程度不高[20]．教学中片面强调知识、技能，忽视数学与现实的联系，也使得学生较少有机会去质疑结果是否符合现实．

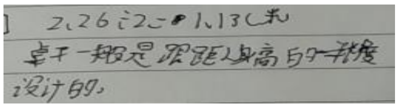


图 5　学生 A 回答例 1 的表现

6　结论及启示

首先，研究以过程为导向构建了双维多水平数学建模能力测评框架，该测评框架包括覆盖广度和覆盖深度两个维度．专家咨询结果显示，专家对研究构建的建模能力测评框架的认同度较高，因此，测评框架具有较好的专家信效度．

其次，研究基于双维多水平数学建模能力测评框架自编了测试题，并进行了一次小规模的调研，以测量中小学生的数学建模能力．结果显示，多位评分者分别依据测评标准对试卷进行评分的一致性较高，说明评价标准具有较好的信度和可操作性．同时，测试结果也表明该测评框架和标准不但能够清晰地体现建模的基本过程，还能够较好区分学生在不同建模阶段的表现水平．

最后，调查结果发现，学生在面对真实情境的数学建模问题时表现较差．在实际教学中，教师可以增设真实情境问题，让学生有机会积累解决数学建模问题的经验，同时教师也可以依据评价标准对学生的数学建模能力进行诊断，并提出针对性的指导意见．Greer 认为数学建模是使用一个或多个数学模型来组织和描述一种情况或现象的过程，是将某种现实情况“数学化”的过程[21]．因此，在设计情境问题时应注意以下几点．(1) 情境应贴合学生的实际生活，让学生有机会结合现实经验理解题目隐含的信息，只有这样学生才能够尝试假设、构建和检验模型．(2) 应合理处理数学内容与问题情境的关系．简单地说，设计的问题应该是具有现实意义和研究价值的“真问题”．(3) 在设置问题难度时，除了预设不同难度的问题，还可以增加一些条件和结果开放的问题，让不同能力水平的学生均有作答机会，以便更好地诊断学生的数学建模能力水平，并提供个性化的教学设计和学习建议．

[参 考 文 献]

[1] 中华人民共和国教育部．普通高中数学课程标准（2017 年版）[M]．北京：人民教育出版社，2018：1．

[2] HENNING H, KEUNE M. Levels of modeling competencies [M] // BLUM W, GALBRAITH P L, HENN H W, et al. Modeling and applications in mathematics education: The 14th ICMI study. New York: Springer, 2007: 225–232.

[3] 朱娅梅．义务教育阶段学生数学建模能力评价框架和行为测评指标[J]．数学教育学报，2018，27（3）：95．
[4] ZÖTTL L, UFER S, REISS K. Assessing modeling competencies using a multidimensional IRT approach [M] // KAISER G, BLUM W, FERRI B R, et al. Proceedings of the 14th international conference on the teaching of mathematical modeling and applications. New York: Springer, 2011: 427–436.
[5] CAI J, CIRILLO M, PELESKO J, et al. Mathematical modeling in school education: Mathematical, cognitive, curricular, instructional and teacher education perspectives [M] // LILJEDAHL P, NICOL C, OESTERLE S, et al. Proceedings of the joint meeting of PME 38 and PME-NA 36, PME-NA. British Columbia: Vancouver, 2014: 145–172.
[6] GREEFRATH G. Using technologies: New possibilities of teaching and learning modeling-overview [M] // KAISER G, BLUM W, FERRI B R, et al. Proceedings of the 14th international conference on the teaching of mathematical modeling and applications. New York: Springer, 2011: 302–303.
[7] FERRI B R. Theoretical and empirical differentiations of phases in the modeling process [J]. ZDM, 2006, 38 (2): 86–95.
[8] GALBRAITH P, STILLMAN G. A framework for identifying student blockages during transitions in the modeling process [J]. ZDM, 2006, 38 (2): 143–162.
[9] OCED. PISA 2021 mathematics framework (second draft) [M]. Paris: OECD Publishing, 2018: 21–22.
[10] MAAB K. What are modeling competencies [J]. ZDM, 2006, 38 (2): 113–142.
[11] VORHOLER K. Measuring metacognitive modeling competencies [M] // STILLMAN G A, BLUM W, KAISER G. Mathematical modeling and applications. New York: Springer, 2017: 175–185.
[12] BICCARD P, WESSELS D. Development of affective modeling competencies in primary school learners [J]. Pythagoras, 2011, 32 (1): 1–9.
[13] BICCARD P, WESSELS D. Six principles to assess modeling abilities of students working in groups [M] // STILLMAN G A, BLUM W, KAISER G. Mathematical modeling and applications. New York: Springer, 2017: 589–599.
[14] 史宁中．数学思想概论：自然界中的数学模型[M]．长春：东北师范大学出版社，2015：104–130．
[15] KAISER G，STILLMAN G. A global survey of international perspectives on modeling in mathematics education [J]. ZDM, 2006, 38 (3): 302–310.
[16] JENSEN H T. Assessing mathematical modeling competency [M] // HAINES C, GALBRAITH P L, BLUM W, et al. Proceedings from the 12th international conference on the teaching of mathematical modeling and applications. UK: Horwood Publishing Limited, 2007: 141–148.
[17] 辛自强，张莉．基于关系—表征复杂性模型的数学应用题表征能力测验[J]．心理发展与教育，2009，25（1）：34–40．
[18] BLUM W, LEIB D. How do students' and teachers deal with modeling problems? [M] // HAINES C, GALBRAITH P, BLUM W, et al. Mathematical modeling: Education, engineering and economics. Chichester: Horwood, 2007: 222–231.
[19] GREER B, VERSCHAFFEL L, MUKHOPADHYAY S. Modeling for life: Mathematics and children's experience [M] // BLUM W, GALBRAITH P L, HENN H W, et al. Modeling and applications in mathematics education: The 14th ICMI study. New York: Springer, 2007: 89–98.
[20] 黄健，鲁小莉，王鸯雨，等．20 世纪以来中国数学课程标准中数学建模内涵的发展[J]．数学教育学报，2019，28（3）：22．
[21] GREER B. Modeling reality in mathematics classrooms: The case of word problems [J]. Learning & Instruction, 1997，7 (4): 293–307.

The Construction of a Two-Dimensional and Muti-Level Evaluation Framework

ZU Dan[1], DING Rui[1], KONG Fan-zhe[2]

(1. Faculty of Education, Northeast Normal University, Jilin Changchun 130024, China;

2. School of Education, South-Central University for Nationalities, Hubei Wuhan 430074, China)

Abstract: Mathematical modeling capacity, as one of the core elements of mathematical capacities, is of great importance to foster and improve students' mathematical key competences. Based on the procedural features of mathematical modeling, this study constructs a two-dimensional and multi-level evaluation framework for mathematical modeling capacity from both vertical and horizontal perspectives. The two dimensions are coverage span and coverage depth respectively. Coverage span focuses on the completion of modeling process, and coverage depth focuses on the level of model assumption ability, model construction ability and model testing ability. Based on SOLO classification theory and the "Relational-Representational Complexity Model", proposed by Xin Zi-qiang, and combined with the analysis of students' performance, the evaluation criteria of each modeling sub-capability under the coverage depth are constructed. The expert consultation and test results confirm that this evaluation framework and evaluation criteria of the two-dimension and multi-level mathematical modeling capacity have high reliability and validity, and can distinguish and evaluate students' modeling capacity as expected.

Key words: mathematical modeling; mathematical modeling capacity; evaluation framework; coverage span; coverage depth

[责任编校：陈隽、陈汉君]