

全球教育展望

10
2018

GLOBAL EDUCATION



中华人民共和国教育部主管
华东师范大学主办
中国人文社会科学核心期刊
中国期刊方阵双效期刊
全国中文核心期刊
CSSCI来源期刊

特约稿

转变教育范式：为何互相借鉴已经不再适用

..... 赵 勇 著 殷丹青 译 宋 玮 校译 (3)

当代课程改革

普通高中课程改革：世界性的课题与经验 崔允漷 冯生尧 (29)

指向学科核心素养的考试命题 杨向东 (39)

深度学习视域下的课堂变革 马云鹏 (52)

基础教育阶段课程统整的误区与矫正

——基于教师视角的质性研究 俞 冰 杨 帆 许庆豫 (64)

传统课程智慧

童子凭什么读“四书”？

——古代“《小学》终，至‘四书’”的课程设计探由

..... 杜成宪 阴崔雪 孙鹏鹏 (77)

国际课程经验

杜威在民国时期受到的批评 丁道勇 (90)

美国保守主义对教育发展的影响

——基于保守主义联盟思想特征的分析 刘正正 高 伟 (102)

儿童中心学校的兴起与美国教育变革 张斌贤 周梦圆 (116)

指向学科核心素养的考试命题

杨向东*

摘要 如何命制指向学科核心素养的考试题目是推进我国当前基础教育课程改革的现实问题。学科核心素养不同于固定的学科知识和技能,关注的是学科课程学习的真实性学业成就。学科核心素养的测评本质上是以建构为核心的评价,需要以学业质量标准建构理论,深刻反思当前双向细目表的不足,构建整合真实情境、课程内容和学科核心素养等多个维度的评价框架。在命制考试题目时,需要准确把握学科核心素养的内涵和外延,发掘特定现实情境用于考查学科核心素养的可能性,基于学科核心素养与情境特征的明确关系确定合理的任务呈现方式,提供恰当的支撑性材料或信息,充分考虑设问指向和设问方式对问题空间大小的影响。以学业质量标准为依据,结合具体评价任务,研制包括评价框架、表现水平和表现样例等成分的等级性评分标准。

关键词 学科核心素养; 学业质量标准; 考试命题; 以建构为核心的评价; 等级性评分标准

作者简介 杨向东/华东师范大学课程与教学研究所兼职研究员、教育心理学系教授(上海 200062)

随着普通高中课程标准的颁布,如何合理测评学科核心素养成为我国基础教育改革和发展的重要议题。虽然学科核心素养的测评会更加关注各种新型评价形式,但在一段时期内,如何变革既有的纸笔考试形式,设计指向学科核心素养的考试题目,是摆在我国教育研究和实践工作者面前的现实问题。

一、学科核心素养测评的性质

学科核心素养的提出,旨在引导对学科育人价值的反思,明确学科育人目标,转变学科学习方式和育人模式。开展学科核心素养的测评,需要深刻理解这一属性的性质,明确其理论依据及要关注的关键问题。

1. 学科核心素养指向学科课程学习的真实性学业成就

学科核心素养是在反思学科本质观的基础上,对学科育人价值的凝练,是学生课程学习中形成的,能够灵活的整合学科观念、思维方式、探究模式和知识

* 本文系华东师范大学2017年度人文社会科学重大预研究项目“普通高中核心素养发展的评价体系研究”的阶段性成果。

体系,应对和解决各种复杂的、不确定的现实生活情境的综合性品质。学科课程的学习过程,本质上是儿童逐渐获得人类积累的文化观念、方法、工具和资源,从经验的与世界互动向理性自觉的认识世界和参与社会的转变过程^[1]。在这个意义上,学科核心素养关注的是学科课程学习的“真实性学业成就(authentic academic achievement)”。真实性学业成就至少要满足三个特征^[2]:(1)能够创造知识(production of knowledge),而不仅是再现他人知识或对他人的知识做出反应。(2)能够开展严谨的探究(disciplined enquiry),深刻理解问题,采用新颖方式整合或重组已有知识,创造新的知识。(3)具有超越评价的审美、实用或个人价值(values beyond assessment),即对个体和社会都有价值的观念、方法、方案或产品。真实性学业成就不只是习得事实性的学科知识和概念,而是能够运用这些知识或概念解决复杂的现实性问题^[3]。它重视不同知识、方法或态度在深层意义上的整合和运用,关注学生在复杂的开放性情境中的综合表现^[4],强调学生在知识应用过程中形成灵活有效的问题解决技能,学会如何与他人合作,如何计划、监控和评估学习方案和进程,学会自主学习和终身学习^[5]。

显然,学科核心素养指向的是复杂的理论建构(theoretical construct),对既有测评理论和实践提出了严峻的挑战。首先,学科核心素养比知识或技能更具有综合性,在具体内涵和表现机制上也更为复杂。其次,学科核心素养是更为高阶的抽象概念,是个体内在的品质或特征。它指向个体在复杂的开放性情境中的外在表现,但其本身并不等同于这种表现。学科核心素养无法直接观测,但可以通过学生在具体任务中的实际表现加以推测。要确保这种推测的合理性,就必须充分了解所测核心素养的内涵和外延,建立核心素养与具体任务上的外在表现之间的关联。

2. 学科核心素养测评是以建构为核心的评价

建构是指评价任务要考查的某种属性。在一般的意义上讲,教育评价是一种基于证据的推理过程^[6],即通过学生在某些评价任务上的表现来推断他们在某些(心理或教育)建构上的特征或水平。评价人员最为关心的不是学生在特定任务上的外在表现,而是导致这些表现背后的各种建构。以建构为核心的评价方式(construct-centered approach)始于对所测建构本身的反思。首先,从教育理念或目标出发,应该测评的是什么样的复杂知识、技能或其他属性?其次,什么样的行为或表现能够揭示这些建构?再次,什么样的任务或情境能够引发这些行为^[7]?

从这个观念出发,学科核心素养测评需要依据如图1所示的测量模型,明确以下内容:(1)所测学科核心素养的内涵和外延、基本构成及其结构关系;(2)这些学科核心素养和课程内容、教学方式以及学生水平之间的依存关系;(3)学科核心素养在不同发展水平上的表现特征及其和学生年龄或学段的关系;(4)结合具体领域,这些素养从低到高、从简单到复杂、从学校学习到现实生活的发展机制;(5)能够有效引发上述素养表现的典型任务情境及其特征;(6)处于不同素养水平上的学生在这些任务情境中的可能表现,这些表现和素养水平之间的内在关系和影响机制等。这样一种建构驱动的素养理论及其

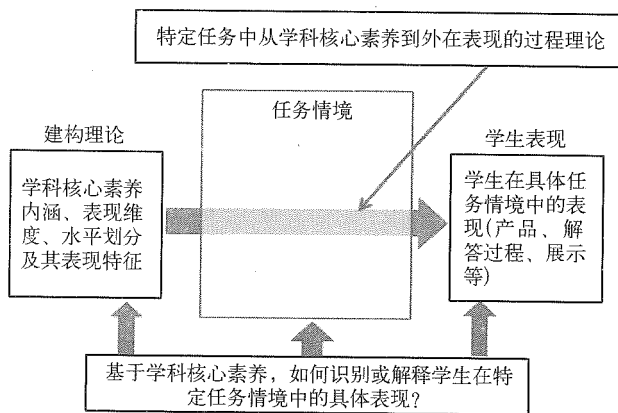


图1 以建构为核心的学科核心素养测评模式

相关证据是学科核心素养测评的理论依据^[8]。不仅如此,该理论还是支撑评价框架制定、评价任务设计、数据收集方案、评价标准研制、数据分析以及结果解释等一系列环节的实质基础^[9]。

3. 学业质量标准即学科核心素养测评的建构理论

修订后的普通高中课程标准中的学业质量标准,是以学科核心素养为纲,整合了特定课程领域的知识、技能、方法或观念,对学生学业成就的整体描述和界定。这种学业质量标准扬弃以学科知识点为纲,以识记、理解和应用为质量水平,转而采用一种整合的、实践取向的学业成就观^[10]。首先,它强调的不是学科固定的知识和技能,而是特定学段结束后学科核心素养表现及其特征。明确界定每个学科核心素养的内涵、外延、构成成分及其结构关系是确立学业质量标准的前提。每个学科核心素养都是在整合或重组学科知识,以学科特有的观念、思维方式、探究模式应对现实情境中形成的。其次,学业质量标准提供了对各学科核心素养不同水平及其表现特征的刻画。与知识技能的掌握不同,核心素养的每个水平都指向现实问题或任务的解决。不同水平之间的差异不是从知到行、从部分到整体的关系,而是知行统一的深度和广度上的差异,体现的是学生在学科核心素养上具有实质意义的质性变化。这种质性变化,除了体现在对特定现实情境或任务的判断和分析,对资源的调动、重组、整合和运用,对方案、进程或社会互动的计划、监控和评估,对结果的解释和论证等,还体现在相关学科知识和技能的结构化程度,学科观念、思维方式和探究模式上的发展水平和综合程度等方面。

素养本位的学业质量标准明确了各学科的育人价值和质量要求,系统阐明了学生在高中阶段素养发展水平及其表现特征。它将学科核心素养视为贯彻基础教育阶段的学生内在品质或特征。像身高、体重刻画物理属性的水平一样,特定学段的学业质量标准刻画了该学段学生在学科核心素养上的水平。在设计学科核心素养测评时,要以学业质量标准为依据,确定考试命题的价值方向、理论框架和水平依据。

二、学科核心素养测评的评价框架

评价框架是测验编制的蓝本。即便有完善的建构理论,研究者仍然可以在不同抽象水平或从不同角度提出各种假设。评价框架通过某种正式的方式(比如语言、图表、模型等),对当前所涉及的建构内涵、内容领域或情境范围、任务类型或特征,以及这些要素之间的关系加以明确界定或限定,并给出具体清晰的表述^[11]。学科核心素养测评在评价框架的理论依据、关注维度和表现形式上都和既有测评有所不同。

1. 既有教育考试的评价框架及其反思

长期以来,国内外教育考试通常采取“双向细目表”的基本形式(见表1)。其中一个维度规定考试涉及的学科知识或技能,另一个维度用于刻画知识或技能上的不同掌握水平。依据该表,评价者确定不同内容或掌握水平上的题目数量、难度以及要采用的题目类型等。这种评价框架的主要依据是布鲁姆的教育目标分类学。依据该分类框架,学科教育教学中的每个目标都可以表述成为教育者所期望的学生认知过程与课程内容(知识)的结合体^{[12][13]}。比如,“学生应该理解变量、表达式和等式等概念”。其中,“理解”这一动词表述的是预期的学生认知过程,“变量、表达式和等式”这些名词则描述了学生应该习得的预期(课程)内容。通过这种方式,学科知识体系被分解为一系列“知识点”,即具体而微的概念、原理或技能,学生在这些知识点上的掌握程度成为预期学习目标。自上世纪80年代传入我国后,该框架在我国基础教育领域得到广泛应用。后来,认知过程逐渐被简化为识记、理解和应用三类,成为确定学生掌握水平的典范术语,也成为我国教育考试命题的理论依据。

表1 双向细目表形式的评价框架

	知识点 1	知识点 2	知识点 3	知识点 4	知识点 5
识记					
理解					
应用					

然而,这种评价框架指向的是一种以学科知识点为纲、以知识点掌握水平为质量水平的学业质量观。它所适应的是以教师讲解为主要教学方式,以学生掌握碎片化知识和孤立技能,在有标准答案的问题或任务中识别和运用为主要目标的教育形态。它过于关注学生在知识点上的认知,缺乏对事物或现象的整体认识和思考,缺少知识整合和综合运用。这种评价框架不适用于学科核心素养的测评。

2. 学科核心素养测评的理论依据

测评学科核心素养需要重新反思学习观,正确理解学科核心素养与学科知识技能之间的关系。学科核心素养所蕴含的学习观是建构主义^{[14][15]}和情

境认知理论^[16]。按照这种理论,知识是个体依据自身经验建构意义的结果。学习是个体在与情境的互动中创生意义的过程。学习应该在具体的、有意义的情境中展开,并且会受到具体任务或问题情境的深刻影响。情境通过活动,并和活动一起共同创生知识^[17]。

这种观点体现了一种多维度整合的学习视角。Deblock(1972)曾将学习概括为如下四个维度的整合:(1)从事实到概念,到关系,再到结构;(2)从事实到方法,到学科方法论,再到学科本质观;(3)从知道到理解,到应用,再到综合;(4)从有限迁移,到中等程度迁移,再到全面的迁移^[18]。其中第一个维度指向学科知识和技能。该观点强调了从事实入手,通过让学生与源于现实世界的真实情境互动,实质性的形成概念和掌握原理,并在此基础上形成结构化的知识和技能。第二个维度即当前所说的学科过程与方法,依然强调从事实入手,让学生在解决情境化任务过程中潜移默化的形成方法,发展思维。不仅如此,学科方法的掌握不能停留在零碎技能层面,需要指导学生从日常经验出发,逐渐过渡到学会用学科的概念、符号、方法和思维来分析情境和提炼问题,解释和论证现象。在这种过程中,学生潜移默化的理解和掌握学科思维方式和探究模式,进而达到能够深刻理解学科的观念及其性质,反思学科的价值和不足。

生活中存在各种不同复杂程度的情境。从简单情境中可以相对容易的提炼特征,形成概念。复杂开放的情境则需要整合相关概念和原理,灵活运用理论或方法才能深刻理解。通过经历不同的现实情境或主题,个体逐渐从运用概念分析和理解情境,逐渐过渡到能够整合不同的学科知识、方法和观念,解释和论证复杂开放的现实情境。这一过程体现的是第三个维度的进程。第四个维度即是这一进程的外在表现,从简单相似的情境到复杂开放的情境,再到跨学科的、整合性的现实生活情境,学生逐渐发展了灵活有效的问题解决策略,培养了自主学习和反思能力。

整合上述四个维度,可以构建如图2所示的学科核心素养发展模型。学生在经历各种真实情境过程中,学科知识和技能不断结构化,学科思维方式、探究模式和价值观念逐渐得以形成,并在应对和解决各种复杂开放的陌生任务时不断得到整合和运用。当学生能够整合已有的结构化知识和技能,运用学科思维和观念开展严谨的探究活动,灵活的、创造性地解决或应对各种复杂现实任务或情境时,就表现出了高水平的素养。

3. 学科核心素养测评的评价框架

按照前面的理解,学科核心素养的测评需要整合多个维度,整体思考学科核心素养、任务情境和课程内容之间的关系。

首先,深刻认识不同复杂程度的真实情境在学科核心素养测评中的重要价值。现实情境不仅仅是学科核心素养形成和发展的途径和方式,也是评价学科核心素养的重要依托。狭义的情境可以是评价任务或教学活动所依存的现实主题、人物、事件及其时空关系。广义的情境即个体所处的日常生活、社会和文化^[19]。评价的功能就在于选择或创设合理的情境,以引发学生在其中

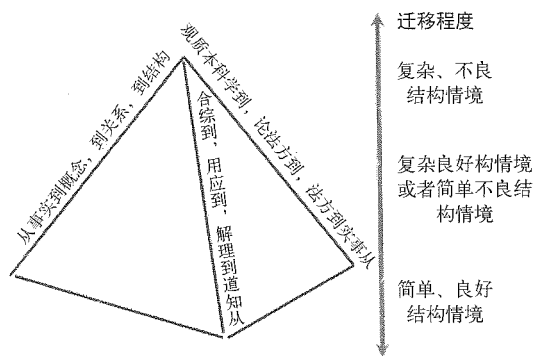


图2 学科核心素养的发展模型

的表现,从而观察和判断他们适应和改变日常实践的水平。

其次,学科核心素养测评不能脱离学科课程内容。系统的、结构化的学科知识、技能、观念和方法,是学生深刻理解任务情境、明确问题、形成假设和问题解决方案的根本基础。大量研究表明,拥有大量的领域知识和问题图式是领域专长的关键特征之一^[20]。专家拥有高度结构化的领域知识,能够将大量信息按照一种有意义的方式组织起来。这使得他们能够超越表面现象,准确把握情境内在特征或问题结构^[21]。因此,学科核心素养的测评指向的不是孤立的、细碎的学科知识点,而要强化学科内容的结构性和关联性,突出在现实情境中思想方法和探究技能的运用。

第三,拥有结构化的学科知识和技能,理解和把握学科思维方法和价值观念,并不意味着就具备了学科核心素养。学科核心素养更多的指向学生在面对不确定的情境时能否做出恰当的反应。当面对复杂问题或陌生情境时,具有学科核心素养的个体知晓如何有效调用和运用已有资源,能够及时反思,自我调节和监控,根据现实情景随时调整策略,重新组织和设计解决方案,更多表现出的一种“适应性专长(adaptive expertise)”的特征^[22]。他们“能够灵活地面对新情境并终其一生持续学习。他们不仅运用他们已经学会的知识,还会运用元认知,持续质疑自己已有专长水平并试图超越它们。他们不仅仅尝试更有效率地做同样的事情,而且尝试着把事情做得更好”^[23]。

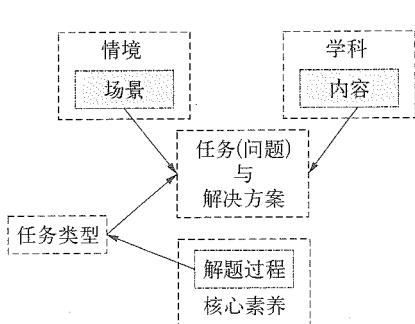


图3 国际学生测评项目(PISA)中的评价框架

因此,学科核心素养测评可以考虑采取如图3所示的国际学生测评项目(PISA)的评价框架^[24]。不同于双向细目表,该框架整合了情境、学科内容和核心素养三个维度。情境提供了联接现实生活世界和学科领域的桥梁,为引发学科核心素养的表现提供了载体。以真实情境为基本素材,通过改变情境的结构化程度、要素数量及其关系,可以创设各种复杂程度的情境化任务。简单的、结构良好的情境可以

考查学科核心素养的较低水平。复杂的、开放性的情境蕴含着大量干扰因素,考查学生对相关知识、技能、思维和观念的创造性整合,体现学科核心素养的较高水平。因此,改变情境开放程度、学科内容容量和问题解决过程,就可以系统设计或控制任务解决过程中的心智要求,形成学科核心素养测评的难度梯度,实现对不同学科核心素养水平的合理评价。

三、指向学科核心素养的考试命题

学科核心素养的测评旨在通过学生在具体任务上的表现,推断其在某个素养上的水平。这一推断过程会比传统测评方式的推理更为复杂,体现在整个推理过程的诸多环节,比如,对所测学科核心素养确切内涵和外延的理解是否准确,创设的情境或任务能否合理引发相关的行为或表现,学生在任务上的表现与相应的学科核心素养水平的匹配是否合理,等等。所有这些都与评价任务的设计有着直接的关系。

1. 准确把握学科核心素养内涵和表现特征

深刻理解所要考查的学科核心素养的具体内涵,是开展学科核心素养命题的关键。修订后的普通高中课程标准给出了每个学科核心素养的内涵界定和主要表现,学业质量标准描述了不同素养水平的表现特征。然而,这些表述仍然是在一般意义上刻画的学科核心素养及其表现特征,是跨越了不同课程内容和具体评价任务的抽象理论。命题时,需要结合学段、学生群体、课程内容范围和当前任务对所考查的学科核心素养及其水平加以具体化,明确在当前任务情境下学科核心素养的实质内涵。例如,高中物理课程标准中对“物理观念”表现水平有如下表述:具有物理观念,能从物理学的视角描述和解释自然现象,能应用物理知识解决实际问题。实际命题时,命题人员则需要在此基础上进一步确定:(1)什么是物理观念?在高中阶段有哪些物理观念?这些物理观念在高中阶段和初中阶段有什么内涵或特征上的区别?(2)和当前所要命制的题目相关的物理观念有哪几个?在当前题目中的具体表现又是什么?(3)如何确定这里所说的“自然现象”的确切内涵?就当前命题而言,能够选择的自然现象的范围与复杂程度是什么?(4)“物理学的视角”是指什么?在当前命制的任务中其确切内涵是什么?只有对这些问题有着清晰的界定和理解,才有可能在情境创设、条件界定、设问等方面有着清晰的指向。

2. 基于真实情境,创设能够引发学科核心素养表现的评价任务

所谓真实情境,是指源于现实世界、贴近学生经验的生活场景。现实生活中的真实情境往往具有极其丰富的信息和特征,蕴含着大量的潜在线索和限制^[25]。真实情境下的任务不像传统测试题目那样具有完整明晰的条件和问题结构,通常也没有固定的答案和解题套路,显得更加真实和自然。它让学生觉得是在解决有现实意义的问题,有助于激发学生参与和投入的兴趣。

选择用于评价的真实情境时,不能只关注当前情境与现实生活中的目标情境的相似性或保真程度^[26],还要关注依托该情境而创设的评价任务在引发学科核心素养表现上的蕴含性^[27]。其实,绝大多数现实问题是嵌套于复杂多变的时空和社会条件中的,很少现实情境及其相关条件能够被完全复制,尤其是在纸笔考试形式下。单纯只是关注情境的外在形式,并不能确保相应任务所指向的问题解决过程和学科核心素养测评的目的是相吻合的。这样讲,并不是排斥真实情境在学科核心素养测评中的重要价值,而是强调要从建构的层面审视当前所测建构和目标情境所需建构的相似程度^[28]。如前所述,学科核心素养指向个体在现实情境中所表现出来的特定领域知识、方法和观念的整合或重组,系统严谨的探究能力以及创新性的问题解决。如果某种情境化任务考查了学生在上述表现,即便没有在严格意义上“重现”对应的现实生活情境,也可以被认为是真实的。

依据这种观念,创设合理的、依托于真实情境的评价任务,需要综合考虑如下几个方面:

(1) 寻找学科核心素养表现与真实情境之间的结合点

如前所述,每门学科都有其对应的现实世界和生活现象。在本质上,每门学科都是人类在长期实践和探索过程中发展起来的一种认识世界或参与社会的独特方式,有其特有的知识体系、思维方式和探究模式,孕育着特定的方法论和价值观。而学科核心素养就是经过学科课程学习后学生逐渐形成的能够整合学科知识、技能、方法、态度和价值观念等应对和解决各种现实情境中表现出来的灵活的、迁移性的品质。从这个意义上讲,各种与该学科相关的真实的现实生活情境提供了引发学生学科核心素养表现的可能性。要命制评价学生学科核心素养的任务,就需要命题人员能够在谙熟学科核心素养的内涵和表现特征的基础上,用一种学科专家的眼光审视和考查各种与该学科核心素养有关的现实生活情境,发现特定现实生活情境与所在学科(知识体系、思维方式、探究模式和价值观念)的关系,挖掘该情境用于考查学科核心素养不同水平的可能性。

图4给出了考查普通高中地理课程标准中“综合思维”这一学科核心素养的任务情境案例^①。4题干描述的是发生在现实生活中的一个实际事件,但从地理学科的角度来看,这是一个和地理有关的现实问题。要解决其中提出的问题,学生需要运用和该问题相关的已有经验、地理知识和原理来深刻理解当前情境,明确问题实质和各种可能的假设,综合影响该问题的各种可能因素,形成一个基于证据的、能够自圆其说的解释和论证方案。它不仅需要学生掌握与该区域及该事件有关的地理基本知识、原理和技能,还需要学生能够将相关的知识、技能和思维方法进行整合,并根据当前任务的特定需求加以灵活运用。这样一

^① 本案例是普通高中课程标准修订过程中,地理学科用于检验学业质量标准的表述及水平划分的科学性和合理性而设计的一个测试题目,特此声明并致谢。题目有所改动。

个现实的情境中,不仅提供了学生开展地理意义上的“综合思维”的可能性,还可以观察到不同水平的“综合思维”这一学科核心素养上的具体表现。

失事之前,马来西亚航空公司的航班 370(简称马航 370)与雷达的最后一次接触点出现在如图所示的吉隆坡西北附近。两个小时之后,飞机最后一次与卫星传送信号的地点出现在图中所示的斜线末端区域。之后飞机就消失了。三个星期之后,马航 370 飞机的残骸在留尼汪岛附近海域被发现。专家当时建议按照图中所示的灰色阴影区域对飞机进行搜救。你觉得当时专家的建议是否合理?请解释并论证为什么?

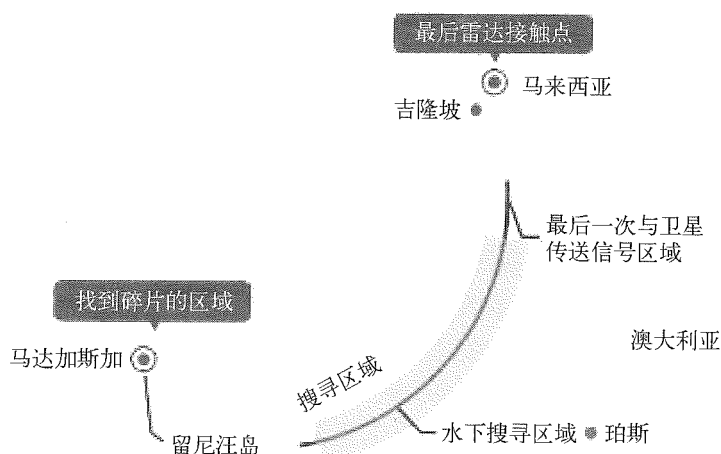


图4 “马航 370 失事”问题

(2) 确定恰当的情境呈现方式

在选择或确定任务情境后,要明确学科核心素养与当前情境的内在关系,从问题解决过程的角度审视情境的呈现或表述方式,建立情境中的不同特征与学科核心素养表现的具体联系,进而确定哪些是关键性的情境特征,哪些属于边缘性的情境特征。所谓关键性的情境特征,是指会影响到情境(或任务)对目标行为或表现的引发的特征。所谓边缘性特征,是对所测建构没有重要影响的情境(或任务)特征,可以在合理的情况下进行适当简化或调整。例如,图 5 给出了同一个情境化任务的两种不同呈现方式。从形式上来看,第二种方式更像是当前我国教育考试中题目的呈现方式,题干采用学科概念和符号进行描述,并且提供了相关变量的确切取值以及背后的理论模型。相比之下,呈现方式一更像是描述真实生活中发生的场景,在表述方式更贴近日常的表达方式。如何确定采取哪一种呈现方式?实际上,从呈现方式一转化成呈现方式二的过程,体现的正是学生能否根据研究问题和情境,运用学科符号系统、概念体系和思维方式,分析模糊的现实情境,从中提炼关键特征或变量,将原有情境表征变成能够揭示其关键本质和内在结构的物理模型的过程。这个过程恰恰体现的是学科核心素养的重要构成。因此,从学科核心素养的表现角度来看,情境的呈

现方式一是更加合理的,而当前考试中常见的呈现方式二反而将这些表现排除在问题解决过程之外,只剩下根据理想模型套用公式和计算的成分了。

呈现方式一

据报道,1980 年一架美国战斗机在威克夫区上空与一只秃鹰相撞,飞机坠毁。试估算秃鹰对飞机的冲撞力。

呈现方式二

一架质量为 M 的飞机与一只质量为 m 的秃鹰在威克夫区上空发生正面碰撞,飞机坠毁。已知秃鹰质量 $m=1$ 千克,身长 20 厘米,飞行速度 $v=30$ 米/秒;飞机质量 $M=100$ 吨,身长 35 米,飞行速度 $V=500$ 米/秒。假设飞机和秃鹰相撞符合完全非弹性碰撞模型,试估算秃鹰对飞机的冲撞力。

图 5 “飞机和秃鹰相撞”问题的呈现方式

(3) 选择适度的支撑性资料或数据

一般而言,学科核心素养测评更多指向整合的、开放性的情境化任务。通常这样的任务是比较复杂的,涉及各种因素和条件。为了让学生更加容易理解题目,命题人员除了采取像图 5 呈现方式二那样更加明确的表述之外,还会以图表等形式提供一些支撑性的资料或数据。但是,需要指出的是,支撑材料或数据提供的方式和数量,会深刻影响到问题解决的空间大小^[29]。因此,在设计基于真实情境的材料题或者以图表形式提供必要信息时,命题者避免因信息提供过多或过少而影响到不同素养水平表现的可能性。解决该问题的一种可能的路径是基于某种理论模型(或框架),建立问题空间与材料或信息提供之间的明确关系。例如,美国国家研究理事会^[30]曾给出科学探究能力的五大关键特征,并针对每个特征列出了不同开放程度的变式(见表 2)。

表 2 科学探究能力的关键特征及不同开放程度的变式^①

关键特征	变 式			
1. 学习者研究科学问题	学习者自己提出问题	学习者从问题中选择,提出新问题	学习者对教师、教材或其他来源提供的问题进行提炼或澄清	学习者研究教师、教材或其他来源所提供的问题
2. 学习者回答问题时突出证据	学习者(自己)决定什么是证据,并进行收集	学习者在指导下收集某些数据	学习者利用提供的数据进行分析	学习者利用提供的数据和分析方法进行分析
3. 学习者根据证据形成解释	学习者通过总结证据形成解释	学习者在指导下根据证据形成解释	学习者按照所给方法使用证据和形成解释	给学习者提供证据
4. 学习者联接解释和科学知识	学习者独立检验其他资源,形成与解释的联接	学习者被引导到某些科学知识领域或来源	向学习者提供可能的联接	

① 根据本文需要,对原表有所改动。

关键特征	变 式			
5. 学习者交流和论证解释	学习者通过合理的、逻辑性的论证来交流解释	学习者在阐述解释的过程中得到指导	向学习者提供改进交流的指导意见	向学习者提供交流的具体步骤和程序
	大-----问题空间-----小			

表 2 中的五大特征涵盖了科学探究能力的不同维度。自左至右学习者得到的资源或信息的支撑逐渐增加,问题的结构化程度越来越高,问题空间则不断减小。然而,任务越是结构化,越是逐渐丧失了观察不同水平的学生科学探究能力表现的机会。

(4) 学会合理的设问

设问不仅影响评价任务所指向的建构的性质,设问方式还会影响到任务的蕴含性。常见教育考试中任务的设问通常指向具体的概念或技能的掌握。学科核心素养的测评则需要改变当前熟悉的设问方式,充分考虑设问的指向和对问题空间大小的影响。图 6 是图 4 所示的“马航 370 失事”任务的另外一种设问方式。和图 4 中的设问不同,图 6 给出了结合“马航 370 失事”题干和图片的三个具体的设问。略加分析就可以看出,问题 1 指向具体的知识点(该海域洋流的性质及方向),问题 2 虽然在题目类型上有别于问题 1,但仍然指向具体的地理知识(该海域洋流对区域气候的影响)。问题 3 看似考查学生基于证据的论证能力,但在问题 1 和 2 以及该题本身的不断提示下,论证所需相关要素或线索已经逐一明确,导致问题空间受到极大的限定,原本指向的论证能力已经简化为相对固定的模式识别。表面上三个设问层层递进,但却将所要考查的“综合思维”降解为零碎的知识和孤立的技能,丧失了观察学生真正的素养表现的机会。

- 问题 1: 流经澳大利亚西海岸附近海域的洋流为
- A. 自南向北流的暖流

B. 自南向北流的寒流

C. 自北向南流的暖流

B. 自北向南流的寒流
- 问题 2: 描述该洋流对澳大利亚西部沿海地区气候的影响。
- 问题 3: 在留尼汪岛发现马航 370 航班飞机碎片后,搜寻专家认为当初确定的搜寻区域是正确的。请绘制该海域大洋环流示意图,以论证搜寻专家的观点。

图 6 “马航 370 失事”问题的另一种设问方式

3. 以学业质量标准为依据,研制等级性评分标准

整合开放的情境化任务给学生提供了展示各种表现、思维方式和问题解决过程的空间,但也带来了如何合理评分的难题。这就需要根据具体的考试题目,开发相应的评分标准。

评分标准应该包括三个成分:评价框架、表现水平和表现样例^[31]。评价框架指向评价任务所要测量的核心素养的关键维度和指标。这些维度和指标要能够真实反映不同水平学生在解决问题过程中所展示的各种结果、表现以及背后的思维特征和探究方式。表现水平是在框架所界定的维度和指标上,不同素养水平学生

在解决任务中所表现出来的具体特征。表现样例是指不同水平学生在该任务上的实际表现。表现样例起到例证评价标准,使描述具体化,易于理解和操作的作用。

研制评分标准可以采用演绎方式,依据学业质量标准引出当前任务的表现标准,将学业质量标准的不同水平表述结合当前任务加以具体化。也可以采用归纳方式,通过分析学生对开放性任务的各种反应,揭示不同个体的关键特征和理解方式,归纳为不同类型或等级,作为当前任务的评分标准。实际研制中可以两者结合,充分利用理论分析和专家经验,在经验做法上开展诸如大声思维等的认知研究,提高评分标准的科学性和合理性。

参考文献:

- [1][10] 杨向东. 基于核心素养的基础教育课程标准研制[J]. 全球教育展望, 2017(10): 34-48.
- [2] Newmann, F. M. & Archbald, D. A. The Nature of Authentic Academic Achievement [A]. Berlak, H., Newmann, F. M., Adams, E., Archbald, D. A., Burgess, T., Ravan, J. & Romberg, T. A. Toward a New Science of Educational Testing and Assessment [C]. Albany, NY: State University of New York Press, 1992.
- [3] Chi, M. T. H., Feltovich, P. & Glaser, R. Categorization and Representation of Physics Problems by Experts and Novices [J]. Cognitive Science, 1981(5): 121-152.
- [4] Cognition and Technology Group at Vanderbilt. The Jasper Project: Lessons in Curriculum, Instruction, Assessment, and Professional Development [M]. Mahwah, NJ: Erlbaum, 1997.
- [5] Hmelo, C. E. Problem-Based Learning: What and How do Students Learn? [J]. Educational and Psychological Review, 2004(3): 235-266.
- [6] Mislevy, R. J. Test Theory Reconceived [J]. Journal of Educational Measurement, 1996(4): 382-396.
- [7] Messick, S. The Interplay of Evidence and Consequences in the Validation of Performance Assessments [J]. Educational Researcher, 1994(2): 13-23.
- [8][11] Mislevy, R. J., Steinberg, L. S. & Almond, R. G. On the Structure of Educational Assessments [J]. Measurement: Interdisciplinary Research and Perspectives, 2003(1): 3-62.
- [9] 杨向东. 理论驱动的心理与教育测量学[M]. 上海: 华东师范大学出版社, 2014.
- [12] Bloom, B. S. & Krathwohl, D. R. Taxonomy of Educational Objectives: The Classification of Educational Goals, by a Committee of College and University Examiners. Handbook I: Cognitive Domain [M]. New York: Longmans, Green, 1956.
- [13][18] Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J. & Wittrock, M. C. A Taxonomy for Learning, Teacher and Assessing — A Revision of Bloom's Taxonomy of Educational Objectives [M]. Boston, MA: Allyn & Bacon, 2001.
- [14] Jonassen, D. H. Objectivism vs. Constructivism: Do We Need a New Philosophical Paradigm? [J]. Educational Technology: Research and Development, 1991(3): 5-14.
- [15] Vygotsky, L. S. Mind in Society: The Development of the Higher Psychological Processes [M]. Cambridge, MA: The Harvard University Press, 1978.
- [16][17][19] Brown, J. S., Collins, A. & Duguid, P. Situated Cognition and the Culture of Learning [J]. Educational Researcher, 1989(1): 32-41.
- [20] Feldon, D. F. The Implication of Research on Expertise for Curriculum and Pedagogy [J]. Educational Psychology Review, 2007(2): 91-110.
- [21] National Research Council. Knowing What Students Know: The Science and Design of Educational

- Assessment [M]. Washington, D. C. : National Academy Press, 2001.
- [22] Hatano, G. & Inagaki, K. Two Courses of Expertise [A]. Stevenson, H. , Azuma, H. & Hakuta, K. Child Development and Education in Japan [C]. New York; W. H. Freeman, 1986: 262 - 272.
- [23] Bransford, J. D. , Brown, A. L. & Cocking, R. R. How People Learn, Brain, Mind, Experience, and School [M]. Washington, D. C. : National Academy Press, 2000: 48.
- [24] Organization for Economic Cooperation and Development. The PISA 2009 Assessment Framework: Key Competencies in Reading, Mathematics and Science [M]. Paris, France: OECD, 2009.
- [25] Lajoie, S. P. A Framework for Authentic Assessment in Mathematics [A]. Romberg, T. A. Reform in School Mathematics and Authentic Assessment [C]. New York, NY: State University of New York Press, 1995: 19 - 37.
- [26] Herrington, J. & Herrington, A. Authentic Assessment and Multimedia: How University Students Respond to a Model of Authentic Assessment [J]. Higher Educational Research & Development, 1998 (3): 305 - 322.
- [27] Greeno, J. G. Gibson's Affordances [J]. Psychological Review, 1994(2): 336 - 342.
- [28] 杨向东. “真实性评价”之辨[J]. 全球教育展望, 2015(5): 36 - 49.
- [29] Newell, A. & Simon, H. A. Human Problem Solving [M]. Englewood Cliffs, NJ: Prentice Hall, 1972.
- [30] National Research Council. Inquiry and the National Science Education Standards [M]. Washington, D. E. : National Academy Press, 2000.
- [31] Lane, S. Performance Assessment: The State of the Art [A]. SCOPE Student Performance Assessment Series [C]. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010.

Item Design for Domain-Specific Key Competence Assessment

YANG Xiangdong

(Department of Educational Psychology, Faculty of Education, East China Normal University, Shanghai, 200062, China)

Abstract: Developing items for assessing domain-specific key competences plays a critical role in moving forward the basic education reform of our nation. Unlike fixed subject-specific knowledge and skills, such competences are regarded as reflecting the authentic academic achievement with which students can creatively solve complex real-world problems by conducting disciplined enquiry and applying integrated domain knowledge. As such, a construct-centered approach can be taken for assessing such competences for which competence-based achievement standard serves as the necessary construct theory. It calls for an innovative assessment framework and new considerations in item design such as situated task development, adequacy of supporting material, ways of questioning, as well as graded scoring rubrics.

Keywords: domain-specific competences; academic achievement standard; item design; construct-centered assessment; graded scoring rubrics

(责任校对: 钟彩凤)