文章编号:1006—9860(2017)06—0059—08

国内外 STEM 教育评估设计的内容分析 *

江丰光[1,2]，蔡瑞衡[1]

（1. 北京师范大学 教育技术学院，北京 100875；2. 北京师范大学 未来教育高精尖创新中心，北京 100875）

摘要：发源于美国的 STEM 教育在近 20 年掀起了全球教育改革的浪潮。国内外教育研究者、教育政策与课程开发等专家和一线教育工作者纷纷开展了 STEM 教育理论与实践的各种探索。随着 STEM 教育的开展，对学生在 STEM 项目的学习和教师的教学如何有效地进行评估成为难题，评估研究也成为近来年教育工作者们关注的重点。该文就 STEM 教育评估设计汇总了 2004 年以来有案例研究的文献并进行整理。通过分析，发现近 5 年来的相关研究数量增长迅速，国内外文献对 STEM 教育领域评估设计的探究是多元而非单一的，主要集中在对学生学习成果的评估。评估设计上应注意完备详实且自成体系，评估方式结合传统和表现型评估，并结合形成性和总结性评估。透过该文的综述分析希冀给国内教育者提供在 STEM 课程与项目的评估设计上参考借鉴。

关键词：STEM 教育；教育评估；内容分析法

中图分类号：G434 **文献标识码：**A

一、STEM 教育发展背景

美国国家科学委员会于 1986 年首次提出“科学、数学、工程和技术教育集成”一说，那时候该概念的缩写是 SMET，被看作 STEM 一词的开端[1]。此后，美国科学基金会把 STEM(Science，Technology，Engineering，Mathematics) 用作描述这四个学科的相关事件、政策、项目和实践等通用的专有名词，于是该名词开始广泛传播[2]。美国政府认为国家能取得的成功就是源于科技创新与发展，而下一代创新能力的关键就是做好教育，尤其是 STEM 教育[3]。如今，STEM 教育已在各个先进国家的教育改革中[4]。各国政府和相关部门也已陆续把 STEM 教育纳入课程体系中。美国研究委员会和国家工程院强调要将 STEM 纳入 K-12 教学，让学生从小接触 STEM 学科知识，为他们在高等教育乃至职业生涯中选择 STEM 相关专业做准备。在这背景下，STEM 教学已陆续出现在各学段的学校中开展，相对应的教学评价也得到关注。目前，如何评估学生在 STEM 课程中能力有所提高是最受关注的议题之一。Ernst 与 Glennie 认为若一味研究改革目标、教导和学习环境而忽视思考对这些目标如何评估，结果是不可行的[5]。随着美国下一代科学标准 (NGSS) 的推出，对 STEM 教育的评估又掀起浪潮。美国纽约州立大学布法罗分校柳秀峰教授将美国下一代科学标准中对于新的课程要求归纳为三个亮点[6]，分别是既强调核心概念，也强调跨学科；从重视科学探究转向重视科学和工程实践，工程教育第一次在世界上认为从 K-12 是必修课程；学习进阶理念，循序渐进，各学科协调，包括语言、社会等。基于上述三个维度，如何评估标准下的学生学习，可得出几点结论：(1) 评估设计要有严谨的系统过程，遵循测量学原理；(2) 综合形成性和总结性来评价学生的学习成果；(3) 课堂内外共同监控与协调评估系统，以提供学生学习的全貌。显然，一套合适的评价工具是观察和判断 STEM 理念是否落实到教学实践的关键[7]。

二、评估的理论基础

评估在教学中占有很重要的角色。Veerle、Dimitri、Pieter、Peter、David 和 Gert 认为，在教学中，评估不仅仅是教学参考的目标，还是贯穿学习过程的目标。评估的模式可以有技巧地融入学习环境中以期更好的学习成果的产出[8]。社会环境对新型人才的要求不局限于对知识的掌握，更重在具

* 本文得到北京师范大学教育学部 2014 年度科研业务费专项资金 (项目编号：CXTD201401) 资助。

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

备能够应用知识解决实际问题的素质。这种强调学生在应对问题时表现出创新技能、信息分析能力和批判思维的培养方式使传统只关注学科知识的评价方式已不适用，近年来 STEM 的研究学者把目光投向了 STEM 的整合实践和评价体系的改革创新。STEM 教育往往在设计上具有整合式和跨学科的特点，在教学上会提供真实的学习环境，呈现出以问题或项目为中心导向的特征，常用到项目式学习(Problem-Based Learning，PBL)。项目式学习通过解决真实生活中的问题，旨在用具有挑战性和富有意义的生活情境问题让学生们参与设计、解决问题、做决策和共同探究的过程，从而锻炼综合能力[9]。这样的课堂和教学聚焦于培养学生处理真实世界问题的能力。因此，STEM 的评估也必须和这样的教学方式相符合。鉴于这种情境模式下的学习，Oberg 认为，评价体系要能够起到指导和扩大高阶学习的作用[10]。Neal 则曾说，一个新式评价体系需要建立在学生表现的基础上，包括一个任务和一张评价单，评价单符合课程目标中对学生绩效质量的要求，其中学生绩效是指学生表现的内容、思考过程以及技能[11]。为了有效评估 STEM 的学习，教师们必须采用和发展一套综合的评价体系，给学生不同的表现机会，从而通过多样的评估方式证明他们的知识储备。综合的评价体系包括了通常所说的三大评价类别：诊断性评价、形成性评价和总结性评价。需要说明的是，虽然是旨在连接学生学习经验和专业知识的真实场景下的学习，需要开发新型的评估方式，但也要把依赖于传统认知评估方式的要求和措施考虑进去，也就是传统认知能力的评估贯穿于整合的以学生表现绩效为评估基础的学习方式中。Ernst 强调将基于传统认知能力的评价方式和基于学生绩效的评价方式两种方式相结合，以进一步提供在教育和专业能力上的评价[12]。还有研究者提议，类似项目式学习这样的新式学习环境应当要有与其教育目标和教学原则相配的评价模式。项目式学习中的学生会通过广泛多样的评价方式被评估。所以综上所述，对于全面整合的 STEM 教学，教学模式是新的，其评估方式也应是创新多元的，在传统关注学科知识评价的同时，更关注怎么评估学生在真实或模拟真实场景下项目式学习的表现和绩效[13]。

三、研究设计

内容分析法作为一种半定量的研究方法，现在被广泛应用于情报学、心理学认知等社会科学领域[14]。它从定性的问题出发，用定量的研究工具和方法对研究对象进行数据分析，最终得到定性的结论[15]。这种方法起源于社会科学采用自然科学的量化分析方法，通常以文献、文本类型资料为研究对象[16]。伴随着 STEM 教育的开展，对应的评估方式在 21 世纪初已有专家关注和报道相关研究。本文研究者以“STEM 教育评估”OR“STEAM 教育评估”“项目式学习评估”为关键词，在 WEB OF SCIENCE 和中国知网上分别搜索从 2004 年到 2016 年的相关文献，通过阅读和筛选，选取可以内容分析的文献。最后得到有具体评估工具设计的文献共 16 篇，国外文章 12 篇，国内文章 4 篇。这 16 篇文章中，有案例研究的文章共 14 篇。我们按作者所在单位的国别将有案例研究的文章分类，发现美国教育学者在相关方面的研究最多，为 7 篇，其次国内包括大陆和台湾共 4 篇，澳大利亚为 3 篇，其他为 2 篇。按出版年份将所有文献分类，可见国外对 STEM 教育评估的研究随着 STEM 教育的关注度有增长的趋势。在最近两年，2015-2016 年对 STEM 教育评估的文献数量增长迅速，2015 年最高，为 5 篇，说明这方面研究的关注度在近两年迅速增长。

四、国内外评估设计种类

笔者根据文献资料，把文献中出现的评估方式按照评估主体、所属类别和出现在不同文章的篇数几个方面整理如表1所示，按不同类型一共区分为 10 种具体评估方式。其中类别包括判断这种评估方式是属于诊断性评估、形成性评估还是总结性评估，以及它属于传统方式 (纸笔测试) 还是新型基于学生表现的评估方式。

表1 评估方式整理

评估方式	评估主体	形成性 / 总结性	传统式 / 表现型	篇数
纸笔测验封闭式题型	教师	总结性评估	传统式评估	6
纸笔测验建构式题型	教师	总结性评估	传统式评估	
学生手稿笔记	教师	形成性评估	表现型评估	3
图示呈现	教师	形成性评估	表现型评估	2
课堂观察	教师	形成性评估	表现型评估	3
作品集	教师	总结性评估	表现型评估	2
访谈面试	教师	总结性评估	表现型评估	1
任务成果展示	师生共同	总结性评估	表现型评估	6
学生自我报告	学生	形成性评估	表现型评估	5
组内互评	学生	形成性评估	表现型评估	2
反思日志	学生	形成性评估	表现型评估	1

注：* 研究者自行整理。

五、STEM 评估案例介绍

我们将整理的有具体 STEM 评估案例的文献内

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

容，按照江丰光副教授所提出的三种 STEM 教育情境分析[17]，认为评估方式主要可以分为四种：正式教育中的教室内 STEM 课程评估、非正式教育下的场馆 STEM 项目评估、学习共同体 STEM 项目评估以及线上 STEM 软技能测试评估。

（一）教室内 STEM 课程

1. 评价设计框架

Routledge 在 2016 年最新出版的 STEM ROAD MAP——一个整合的 STEM 教育的框架一书中，详细阐述了数据驱动下的 STEM 评估量表如何开发。Tony、Kristin、Koskey、Gregory 和 Erin 强烈支持开发一套符合 STEM 教育的综合性评价体系。评价系统是对学生 STEM 学习中对所要达到目标的程度测量和界定，在任何 STEM 课堂中，一个可靠的评价体系都必须是全方位的，这是因为引导课堂的学习目标本身就是多维度的[18]。STEM 课程和教学策略逐渐凸显出跨越学科的特性，并与真实世界越来越紧密相关。要做到成功评估 STEM 的学习，其评价体系需不断优化，提高效果和效率同时减小误差的可能。为此，他们提出第一步要将学习目标量化，并明确定义认知的不同水平，从记忆理解的低水平到分析应用的高水平都需定义。学习目标如何定义要关注学生的行为表现，并且每个操作化定义的学习目标都都需明确，就是聚焦在一点。其二，他们建议用前后测来判断学生先前的知识水平和学习的进步程度。这说明，传统测试方式在某些方面，尤其是知识概念掌握上还是得到了比较好的认可。对于实践层面上如何评估，Tony 等人概括了评估量表的两种呈现方式：目标型和自我建构型。目标型的题目有一个明确的答案，包括选择题、连线题、判断题，也包括填空题。自我建构题型则要定义量化的等级，一般包括论述、论文、简答、作品集、实践任务以及项目绩效评估。自我建构题型的判分往往涉及到多个相关者，评价主体多元而非单一。

目标型题目因其答案非是即否，不需赘述怎么评判。就自我建构题型而言，笔者整理了一些案例，研究多元主体模式下的评估在具体实践案例中是怎样开展的。

2. 评价阶梯

CAROL、ELIZABETH 和 MERCEDES 在 2004 年提出了一个“评价阶梯”的概念。他们认为，当课堂教学开始出现项目式学习、任务表现这样的特点时，所用到的评价方式应当是一步一步的递进过程，也就是一个阶梯式评价方法[19]。

这是一种挑战传统的新的真实的评价技巧。教师在熟悉使用了前一阶段的评价方式后，方可进入下一阶段评价方式的使用。第一层起步阶段，适用于幼儿园和小学的总结性评价，评估方式可采用常用的测验、作品集和项目列表这几种方法。第二层提高阶段可以作为形成性评价补充第一阶段的总结性评价，包括学习笔记，书面描述，学习图式和自我评价。学习笔记可以帮助老师洞察学生的认知过程和对学习内容的态度。书面描述，老师可以在课程结束时让学生以“我学到了……”的句子为首写下本课程学到的东西，这帮助老师了解学生是如何理解概念的。对于不擅于用语言描述的学生则可以让他们用图式说明，老师便可以从中了解到学生对概念认识的深度。这个过程利于学生批判性思维和创造性思维的形成。此外自我汇报、现场记录和现场观察的方法都有助于了解学生的学习过程。通过自我汇报的方式，老师可以让学生对自己的实验或作品进行评价，总结优缺点。第三层顶峰阶段，让学生自己选择通过任一媒介去陈述表达对概念或客体的自我理解和相关知识。这种方法可以描述为“我会教”策略。CAROL 等人认为这是一种非常好的评价方法，学生必须研究和组织所有学到的信息去教授别人。当然教师本人可以对自我教学进行反思，考虑如何应用评估策略，还可以探究更新颖的评价方式。

可以看到，这个时期 CAROL 等人已经开始挑战传统课堂的评估方式，并且给教师建议了不同水平的具体评价模式，比较综合地关注了学生在科学概念、书面表达、思维过程、自我评价和口头陈述等不同方面的能力。

3. 多元主体模式下的评价

2006 年，Veerle、Dimitri、Pieter、Peter、David and Gert 所做的研究具体说明了高校中的 STEM 教育在多元主体模式下的评价如何进行。他们认为，项目式学习作为符合社会对高等高校毕业生的期待的一种培养学生主动建构知识与培养创新能力的学习方式，也是 STEM 教育中最常见和重要的教学模式[20]。相应地，评估方式的发展也从传统纸笔测验转向新式的评估方法。新式评估方式则提到表现绩效评估，包括作品集评估。从评价主体分类，评估则可分为自我评价和同伴评价。

通过对比利时某高校为时 3 个学期的一门“学生研究”课程的追踪研究，Veerle 等人尝试 5 种不同的新式评价方式并得到了在师生中的反馈。这 5 种新式评价方式按评价主体分类分别是：自我评价，同伴评价，共同（合作）评价，表现评价和反思日志。其中，自我评价和同伴评价常作为过程性评估，共同评价和表现评价常作为总结性评价。所谓自我评

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

价，是指学生自己对学习过程的判断，尤其是过程的产出和成果的相关表现。同伴评价，一般是指组内个人评定其他成员的方法。共同评价就是师生都要进行评价，学习成果收到整合的评判。而表现评价则是指，观察学生是否积极完成复杂、重要的任务，评判他们用所学知识、技能去解决真实问题的表现，以及最终产出成品的效果。自我反思日志，是一种学生以日志的形式总结记录自己学习过程以及学习中的反思的评价模式。这种评价方式并不同于自我评价，它是以随意不拘束的方式自由向老师表达学习想法的过程，这个过程旨在让学生反思自己的学习。反思日志的方式不强迫学生，学生不想做的时候便可以选择不做。研究表明，这五种方式中，作为过程性评价的自我评价和同伴评价，与作为总结性评价的共同评价和表现评价各有优缺点，可以互补，在教育应用中也可以结合使用。而反思日志因其可以获得课程以外的很多沟通信息，促进师生相互的理解和交流而收到最好的反馈。

Veerle 等人的研究把研究对象聚焦在高校，研究的突出贡献在于从多元主体的模式下将适用于项目式评估的不同新型评估方式作了界定和对比，并从实践中得到评价方式在师生中的反馈。作为一篇定性研究文章，对教师选取评估方式有一定的参考价值。

4. 独立任务指标评估量表

Freda and Koplin 定义基于表现的评估有两个关键的特征：一个结构化或开放式的表现型任务，以及一个附带的评价单，评价单上的项目即课程目标中定义的表现的构成要素[21]。这响应了 Neal 的想法。北卡罗来纳州自 2007 年建立了一些 STEM 高中，2014 年这些高中进行重新设计，尝试转向 STEM 学习模式，为了调查学习的效果，Freda and Koplin 在表现绩效评估的理论基础上建构了独立分类指标的任务度量表，并进行了试点研究。评估先根据课程蓝图划定目标，制定表现型任务，教师收集学生完成任务过程的设计手稿，然后对应具体指标量表进行评测。研究中，制定的表现型任务划分结构，评估时，按教学的主题，每个主题有大约 3 个任务，每个任务有大约 5 个具体实施操作，学生要完成的则是这些子任务，也就是共需完成约 15 个小练习。任务的完成会要求在专门的手册上作活动记录。评估量表便是根据记录手册对每个子任务进行不同等级的评分。评分有 4 个分数等级，分别对应：尚未或将要符合标准、基本达到标准、完成目标和出色完成目标。每个子任务对应的每个等级都有详细的阐述，怎样的操作和记录达到该等级有详细说明。

这种评估方式有可行性，但缺点是，学生自己的操作学习可能会因为学习环境和一些不可控因素的影响导致学习效果有差距。由于考察的材料只有学生的作业手册，我们认为这样的方式比较适用于学生有书面记录或手稿的表现型设计任务活动。

此外，Lyn D English 和 Donna T King 针对四年级学生在航空航天方面 STEM 学习的评估做了研究。课程要求学生学习、制作和测试设计一个可以飞的纸飞机，采用的评估方式是从学生对“飞机”的绘画、计算制作和重设计三个维度评估。与 Freda and Koplin 的案例不同的是，评估时，三个维度即分别对应设计能力的三个水平，通过计算达到不同设计水平的学生数来评估整体上学生的设计能力如何[22]。

5. 多种评估方式结合应用于课堂

多种评估方式的结合应用还是比较受欢迎。Hudson、English、Dawes、King & Baker 通过收集档案记录、学生工作样本、学生作品的照片、研究者笔记、教师记录和访谈以及学生书面反馈的资料来调查教师的教学知识实践与小学生在 STEM 教育单元学习成果的关联。学生的产出成果包括作品样本、设计图稿以及在活动中的参与度。对学生 STEM 学习成果的评估集中在学生的作品上，作为补充的还有学生的报告，学生对于作品的注释以及学习手册上对于设计过程中任何方面的记录。设计图稿即在学习手册上。除此之外，在设计作品之前为了考察学生对于材料的认知，教师使用了预测—观察—解释 (POE) 的学习活动模式。对材料属性的认知也会记录在学生手册上[23]。

ISMAIL MARULCU 研究了在结合乐高的基础上，工程导向的课程对比探究式课程对五年级学生学习简单机械的影响研究中，除了采用前后测纸笔测试评测学生的 STEM 基本知识和能力外，还采用面试的形式随机选取学生进行能力考察。他的面试问题清单上陈列了 5 道题，都是紧扣学习内容——简单机械。比如，给学生一个有不同尺寸齿轮的平台，让学生回答再给定的条件下，某一个齿轮将旋向哪个方向，为什么。面试的时候现场备有相应的器材以备使用。面试的时候，老师可以根据学生的回答反应，跟进询问探讨，以试图更好地了解学生的想法。通过这种现场口头表达解决问题的方式，便于评估学生的思维能力和对学习内容的掌握水平 [24]。面试问题的评估指标则也是按不同的作答情况，给予 0–3 分四个不同分数级别。

面试口述的方法比较新颖，笔者认为在需要深入了解学生想法、考察学生设计思维的项目中可以

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

应用。此外，对于小组合作学习或表达交流上不擅长的学生，教师可以用这种方法来辅助了解学生的学情。

6. 关注工程教育的评估模式

工程教育由于相对科学、数学教育起步较晚，并且与生活实际关系密切，是 STEM 教育中备受关注的一环。接受工程教育的高校毕业生要求具备项目设计的能力，掌握持续设计必要的概念知识[25]。在研究基于学习环的持续性模式对学生理解概念的持续性的影响中，Mary、Joshua、Caroline、Michael 阐述了工程教育中如何测试学生所获知识的可持续性。这两种方法分别是概念图和自我报告调查。学生自我报告调查是一种间接测量学生知识可持续性的措施，调查通常要求学生在知识持续性的各个方面做反馈，并使用给定模式作答。学生的表述往往因个人经验在多方面有所不同。这种方法的缺点在于，学生对自己认知理解的感知可能会大于他们实际获得的知识。概念图则是一种新颖的组织知识的图形化工具，用于直接测量学生的认知能力。它可以帮助学生回忆和联系所学的知识，适用于知识的整合和巩固，评价也是从一个整体的框图而不是单个的因素分析，让学生的概念知识得以丰富和饱满。概念图具体怎么评分，Mary 等人认为可以将能力分为 1、2、3 三个量化等级，分别从概念图整体的综合性 (广度、深度)、组织性以及准确性三个维度评判。

7.STEM 教育课程大纲的评价

除了对师生教学和学习效果的评价，也有研究者针对 STEM 教育的课程大纲开发评价量表。土耳其肯特大学的 M. Sencer CORLU 为了调查高校中得到工程技术鉴定委员会 (ABET) 认可的课程是否比未得到其认可的课程在实施过程中更符合 STEM 教学实践，开发了一个针对 STEM 课程大纲设计的评估量表。量表对 STEM 学习师生共同体、STEM 课程整合设计和 STEM 课程中的学习评价三个方面进行评估，评分同样分为 1–4 分 4 个等级，每个方面都划分了 5 个维度。其中，对 STEM 学习师生共同体的评估包括：课外学习经验的可获得性如何、有无以学习为中心的政策和期望、教师对学生的关联如何、学生对教师的关联如何以及学生之间的关联如何；对 STEM 课程整合设计的评估包括：有无深入知识内容的框架、有无 STEM 学习的信念、态度和价值观、有无 STEM 学习技能、有无连接终身学习以及有无整合的和跨学科的知识建构；最为本文关注的是对课程评价方式的评估，包括：是否关注成绩、有无形成性评价、有无总结性评价、有无真实评估和有无自我调节环节。真实评估指贴近生活，有技能要求[26]。自我调节包括了自我评价和同伴评价，学生自己有真正参与在设计过程中。

8. 国内对 STEM 教育评价的应用

对国内的文章以“STEM 教育评估”“STEM 教育评价”和“项目式学习”为关键词进行检索，没有得到关于专门研究开发 STEM 课程评价的文献。不过有对国外相关研究和标准体系的翻译和引进，以及对国外 STEM 教育评估动态的关注。华东师范大学王玲玲做了基于 STEM 的小学科学课程设计研究 ，也是比较典型的案例。王玲玲在分析了三个国外的课程案例后，集合国内教材做了课程设计。三个国外案例分别是，国际技术和工程教育者协会 (ITEEA) 开发的基于设计的工程设计课程项目 EbDTM 项目、由美国科学博物馆创建的 EiE 项目、以及由 SAE International 开发的 AWIM 项目进行分析[27]。王玲玲发现这三个案例中教学评价的方式互不相同。第一个案例采用了是非判断题、简单的建构性回答题与口头陈述相结合的方式。第二个案例涉及整个工程设计的过程，评价是其中一环，让学生反思设计过程。第三个案例的教学模式是将工程设计与探究学习 5E 模式整合形成 6E 教学模式，其中评估和增添的重新设计过程即是学生对自己设计的反思。第二、三个案例用到则是表现评价和学生自我评价。作者总结 STEM 教育的教学评价是多元化的，并且嵌入到教学设计的整个过程，同时具有可调节性。对于国内的课程设计，王玲玲使用了 QIEIE 的模型，即问题—探究—设计与建造—改善—评价五环模式。对学生的学习成果，这里的评价部分则要求学生以小组表达的方式呈现最终方案，并可用不同方式说明选择该方案的原因。评价标准由师生共同建构。评价表由小组成员参与度评价表和最终展示评价组成。展示评价表从展示内容、对主题思考与交流的体现、呈现方式和表达效果 4 个维度展开，每个维度的评分划分为 0、5、10、15 分 4 个等级，不同等级有详细说明。

对于这个案例笔者认为，整个活动过程都是对学生表现的观察与评估，其中学生自己设计与建造和改善作品更是他们自我评价的体现。让学生共同参与建构评价标准这一点的设计尤其值得参考。同时，学生事先已参与设计评价标准，后期在学习实践过程中，会不会有意向或有目的地向评价标准里的项目靠拢从而影响实际学习水平？这将是这种评价方式值得反思的地方。

(二) 非正式场馆 STEM 教育

场馆中利用展示展品帮助学生们在 STEM 学科

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

上的学习也很常见。一般除了评估学生对展品和情感态度外，还可以检测学生参观后的学习感受。台湾国立科学工艺博物馆在2007年和2010年分别举办过“探索90%的未知——丁肇中与基本粒子特展”和“青春宅急便”巡回展。“丁肇中特展”通过随机选取参观观众进行追踪观察和参观结束后访谈的方式调查观众对特展的感受和对其中科学知识的了解[28]。“青春宅急便”巡回展选取特定不同区域主题高中和职高学生为对象，采用调查问卷的方式，包括封闭式(选择题)与开放题，研究他们在参观前和参观后对展品知识的认识[29]。

(三)学习共同体概念STEM项目

亚利桑那州立大学的水资源项目是一个典型的STEM学习共同体项目，该项目以环境与水为主题培养学生21世纪的核心素养，得到来自学校、企业、政府机构以及家庭、社区的支持参与[30]。一方面旨在通过水资源主题的学习，提高学生生活与职业素养、学习与创新素养和信息媒介与信息技术素养等核心素养，另一方面旨在对教师进行专业培训，促进教师专业发展，通过设计相关课程及评价措施改革教学实践。在这个学习共同体项目中，亚利桑那州立大学采用的评价方式则是让学生提交水资源学习的调研分析报告，同时参加研讨会，展示自己的研究成果。评价方式融合了不同的表现型评估。

(四)线上STEM教育评估方式

学校和教师以外，还有研究者开发了更多的课外综合评估方法。对于难于用传统测试方式测量的“软技能”，包括利用信息技术搜集信息，创造力和批判性思维，学习如何思考和解决问题，理解和尊重他人并学会与他人沟通合作，理解不同的文化、信仰、价值观和理念。这些能力可以归为现今人们所说的“21世纪新技能”。软技能如何评估？专家们首先根据已有的国际学生能力评估计划等测试肯定了“软技能”是可以评估的，维多利亚州的课程评估局已经选取其中重要的四个技能，对其详尽设计了技能发展规划，也是按基础水平到十年级划分了水平等级[31]。此外，他们开发了在线测试，目前已经可以用于某些技能的评估。该州米歇尔研究所也在进行一项关于“软能力”评估的试验。

线上测试和交流其实已经比较常见。大量的在线课程提供了线上交流的机会。而对于STEM教育和学习，有一些科学场馆的网站和专门的研究团队在开发线上测试。比较典型的有美国ACT Aspire在线评估系统[32]。ACT Aspire在线评估系统是专为K–12年级STEM教育做的新型在线评估方式，可以为学习者提供学习、评估、课堂和交互的机会。基于课程标准的ACT Aspire整合在线系统常用作阶段性形成性评价，帮助教师、学生、家长以及学校管理者互相促进成长，及时的反馈使教育者可以调整教学和策略。这种方式可以弥补照顾不到有些不喜欢现实中面对面交流学习的学生的不足，常以操作性和游戏式为特点。

六、结论与建议

通过国内外评估文献的分析，我们可发现自2015年以来，世界范围内的相关实践研究增长迅速，评估设计开始趋向系统化。对STEM教育的评估设计逐渐受到重视，并体现出多元评估设计导向。综合的STEM评估体系应当具备这样几个特征：(1)多元主体共同进行。学生能力的考察应当客观判断，不能由教师单方面根据作业作答或课堂表现来判定。(2)不同维度尽可能全面考察。这一点是因为STEM教育目标就是培养能够在不同情景下，整合应用不同知识来综合设计解决真实问题的跨学科人才。STEM的教育活动设计就是综合的，评估也应当从多方面考察，肯定学生表现出色的方面。(3)独立任务指标的评估量表是比较典型的STEM课堂学习或者项目式学习的评估方式。教育者常常愿意将学习目标分解，也就是之前提到的量化定义，然后用一个综合的任务活动或者试题考察学生，任务中不同的操作要求或题目的不同作答要求分别对应不同的学习目标维度，评估时，教师按照学生在不同维度的完成情况划分该学生在各个维度的能力水平。(4)教师在选择评估方式的时候，可以根据学习目标、教学方法、学生特点有针对性地选取，在一个项目中评价方式不一定越多越好，但是可以针对不同维度结合采用不同的评估方式。

每种评估方式各有其优点与限制，为了更好地对比，我们研究整理11种评估方式并以优缺分析如表2所示。各教育工作者可以根据STEM课程设计与教学实践情况采用评估方式。

表2 不同评估方式的优势和限制

评估方式	评估主体	类型	优势	限制	适用范围
1. 纸笔测验封闭式题型	教师	总结性评估	了解学生对知识的掌握	不能了解学生对知识的实际应用	适用阶段性定量评估
2. 纸笔测验建构式题型	教师	总结性评估	了解学生对解决问题的综合思考	不能了解学生的动手实践能力	适用阶段性定量评估
3. 学生手稿笔记	教师	形成性评估	了解学生的构思设计过程	看不到最终学习成产出成果	适用课堂上定性评估
4. 图示呈现	教师	形成性评估	清晰直观地了解学生对知识的掌握，锻炼学生对认知的整合表达	一般只适用与对概念型知识的掌握	适用课堂上定性评估

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

续表 2

5. 课堂观察	教师	形成性评估	观察学生学习过程与教学不足	看不到学习产出成果	适用课堂上定性评估
6. 作品集	教师	总结性评估	了解学生自我设计能力	不能了解学生学习过程中反映的问题	适用阶段性定性评估
7. 访谈面试	教师	总结性评估	可以追踪学生的思考过程，促进师生交流	注意记录的完整性	适用阶段性定性和定量评估
8. 任务成果展示	师生共同	总结性评估	多为小组合作。了解学生综合能力，对知识和技能的应用，对问题解决和产品设计的整体表现；促进表达展示能力，促进小组沟通	不能了解学生的学习过程和组内个人表现	适用阶段性定性和定量评估
9. 学生自我报告	学生	形成性评估	帮助学生自己反思学习过程和效果	学生的自我评估可能不够全面客观	适用阶段性定性评估
10. 组内互评	学生	形成性评估	了解学生在同伴交流过程中的表现	组内同伴的合作度可能会影响评估的客观性	适用阶段性和课堂上定性评估
11. 反思日志	学生	形成性评估	了解学生对教师和课程学习的想法，增进师生情感交流	对于文字表达能力薄弱的学生反思日志撰写较困难	适用课堂外定性评估

注：*研究者自行整理。

根据评估方式的特点最后提出以下几点综合性的建议，希望可以供国内 STEM 教学中的评估设计参考。

第一，操作化定义学习目标。从我们搜集到的政策文件和案例研究文献中可以看到，要想真正测量评估学生在 STEM 这样跨学科整合式的学习中的学习效果，教师和指导者需要制定一个完备详实的评估标准。那么就需要将学生抽象的学习表现操作化定义，便于定量分析。在一个量表中，首要的是要将学习目标，或者说对学生的行为期望分解具体化。

第二，明确要考察学生什么能力以及能力水平等级的划分。正如分析的文献内容所体现，期望学生展示出来的表现行为被操作化定义后，学生真实的学习行为表现也需要有明确的水平等级划分。这样做，一是为了在形成性评估中，教师可以了解学生在学习阶段中对知识、能力等方面的掌握正处于什么程度，二是为了在总结性评估中，教师可以比较直观地看到多少学生达到了哪个程度。明确的水平等级划分，帮助老师更清晰的认识到学生的学习效果和反思教师自己的教学成果。

第三，不同方式结合应用。没有哪一种评估方式囊括所有优点，STEM 教育是综合的，其对应的评估也是系统的整合的。在实际教学中，学生的表现更不能用哪一种方式判断好与坏，需要形成性评估和总结性评估结合观察。如果只用形成性评估，教师可以看到学生在学习过程中的表现，但是看不到学生最后产出的成果，不能给学生一个总体的客观成绩，就无法清楚地定位学生的水平和自己的教学结果。如果只用总结性评估，教师看不到学生的学习进步的过程，无法了解学生是怎样学习的，不知道学生在学习过程中在哪些方面学得更轻松哪些方面学的比较困难，就无法反思自己的教学在哪方面需要改进，更对小组作业中个人的表现不能定位。同理，如果只用传统评估方式，则对学生的实际应用能力发展没有帮助，也不能客观地看到学生多方面能力的发展；如果只用表现型评估方式，则可能忽略学生对基本知识内容掌握的扎实程度。

此外，学生作为学习主体，其实可以参与评估体系的建构。这样教育者与受教育者的相互交流，或许可以促进双方对教育过程的认识，帮助双方对学习目标有不同的认识。总之，评估主体要多元化，评估方式要多样化，评估应当成体系。

最后，给学生自我评估和调整的机会。如同真实的工程设计过程，学习总是在不断进步，作品总是在不断完善。在 STEM 教育的任务式学习中，应当给学生自己反思和重新设计的机会。这不仅可以促使学生对自己的成果思考和评价，还可能有助于学生了解真实的设计流程，提高设计能力。学生的自我调整鼓励学生相互的对比讨论，交流不同意见，鼓励学生共批判的眼光看待问题。

参考文献：

[1] [3]余胜泉, 胡翔. STEM 教育理念与跨学科整合模式[J]. 开放教育研究, 2015, 21(4): 13–22.

[2]钟柏昌, 张丽芳. 美国 STEM 教育变革中“变革方程”的作用及其启示[J]. 中国电化教育, 2014,(4): 18–24.

[4] [5] [13]Ernst J V, Glennie E. Redesigned High Schools for Transformed STEM Learning: Performance Assessment Pilot Outcome[J]. Journal of STEM Education: Innovations and Research, 2015, 16(4): 27–35.

[6] NGSS Lead States. Next generation science standards: For states, by states[M]. Washington, DC; National Academies Press, 2013.

[7]Harwell M, Moreno M, Phillips A, et al. A Study of STEM Assessments in Engineering, Science, and Mathematics for Elementary and Middle School Students[J]. School Science and Mathematics, 2015, 115(2): 66–74.

[8][9][20]Van den Bergh V, Mortelmans D, Spooren P, et al. New assessment modes within project–based education–the stakeholders[J]. Studies in educational evaluation, 2006, 32(4): 345–368.

[10]Oberg C. Guiding Classroom Instruction through Performance Assessment[J]. Journal of Case Studies in Accreditation and Assessment, 2010, (1):1–11.

[11]Neal J. The Power of Performance Based Assessment at the Post–Secondary Level[J]. International Journal of Learning, 2009, 16(9):87–101.

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[12]Ernst J V. Analysis of cognitive and performance assessments in an Engineering/Technical Graphics Curriculum[J]. Analysis, 2008, 45(1):88-99.
[14]孙瑞英.从定性, 定量到内容分析法——图书, 情报领域研究方法探讨[J]. 现代情报, 2005, 25(1): 2-6.
[15]Liang Z, Xianxue M. 文献计量法与内容分析法比较研究[J].图书馆工作与研究, 2013, 1(6): 64-66.
[16]邱均平,邹菲. 关于内容分析法的研究[J]. 中国图书馆学报, 2004, (2): 12-17.
[17]江丰光. 连接正式与非正式学习的STEM教育——第四届STEM国际教育大会述评[J]. 电化教育研究,2017,(2):53-61.
[18] Johnson C C.; Peters-Burton E; Moore T J.STEM Road Map: A Framework for Integrated STEM Education[M]. New York: Routledge, 2015.171-173.
[19]Corcoran C A, Dershimer E L, Tichenor M S. A teacher's guide to alternative assessment: Taking the first steps[J]. The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 2004, 77(5): 213-218.
[21]Hui F, Koplin M. The implementation of authentic activities for learning: A case study in finance education[J]. E-Journal of Business Education & Scholarship of Teaching, 2011, 5(1): 59-72.
[22]English L D, King D T. STEM learning through engineering design: fourth-grade students' investigations in aerospace[J]. International Journal of STEM Education, 2015, 2(1): 14-31.
[23]Hudson P, English L, Dawes L, et al. Exploring Links between Pedagogical KnowledgePractices and Student Outcomes in STEM Education for Primary Schools[J]. Australian Journal of Teacher Education, 2015, 40(6): 134-141.
[24]Marulcu I. Investigating the impact of a LEGO-based, engineering-oriented curriculum compared to an inquiry-based curriculum on fifth graders' content learning of simple machines[D]. Boston:Boston College, 2010.
[25]Watson M K, Pelkey J, Noyes C, et al. Assessing impacts of a learning-cycle-based module on students' conceptual sustainability knowledge using concept maps and surveys[J]. Journal of Cleaner Production, 2016, (133): 544-556.
[26]Corlu M S. Insights into STEM Education Praxis: An Assessment Scheme for Course Syllabi[J]. Educational Sciences: Theory and Practice, 2013, 13(4): 2477-2485.
[27]王玲玲.基于STEM 的小学科学课程设计研究[D].上海:华东师范大学, 2015.
[28]陈玫岑,张美珍.科学展示中的漫画印象[J]. 科技博物, 2007, 11(1&2): 31-45.
[29]苏芳仪,蔡佳燕. "青春宅急便"巡回展学习成效个案分析[J]. 科技博物, 2010, 14(1): 67-90.
[30]江丰光,陈慧. 国际核心素养教育的典型案例分析与启示[J]. 中小学信息技术教育, 2016,(9): 10-14.
[31]THE CONVERSATION: Schools will teach 'soft skills' from 2017, but assessing them presents a challenge[EB/OL].https://theconversation.com/schools-will-teach-soft-skills-from-2017-but-assessing-them-presents-a-challenge-68749,2016-11-21.
[32]Rich C H, & Matovinovic D. ACT aspire online assessment system designed to provide insights to inform K-12 STEM education[C]. Beijing: Beijing Normal University, 2016.

作者简介：

江丰光：博士，副教授，硕士生导师，研究方向为 1:1 学习、学习科学、STEM 教育、信息技术于课堂创新教学、学习空间 (fkchiang@bnu.edu.cn)。

Content Analysis on the Assessment Design of STEM Education at Home and Abroad

Chiang Fengkuang[1,2], CaiRuiheng[1]

(1.School of Educational Technology, Beijing Normal University, Beijing 100875; 2.Beijing Advanced Innovation Center for Future Education, Beijing Normal University, Beijing 100875)

Abstract: STEM Education originates from America, which has set off a great wave of education reform worldwide. The researchers in education, the experts in educational policy and curriculum as well as the front-line educators have carried out a variety of explorations in theory and practice of STEM education. With the spread of STEM education, how to assess students' learning and teachers' teaching in STEM projects becomes a focus. The thesis has used content analysis to collect and summarize literature with case study on assessment design in STEM education at home and abroad since 2004. It is discovered that in these articles, the explored methods of assessments in STEM education are diversified with a focus on students' outcomes. Assessment designing should be complete, detailed and systematic, combining traditional and performance evaluation as well as formative and summative assessment. We hope that the thesis could provide advice on assessment designing on STEM curriculum and projects to educators at home through our literature review and analysis.

Keywords: STEM Education; Assessment in Education; Content Analysis

收稿日期：2017 年 2 月 4 日
责任编辑：赵云建

?1994-2017 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net