

元评估视角下高校创新创业教育评价现状及展望

——以16份评价文本为例

陈佳巍

(湖南大学教育科学研究院, 湖南长沙, 410082)

[摘要] 创新创业教育质量的提高须以科学评价为基础。借助斯塔弗尔比姆 CIPP 评价模式, 从评价设计、评价过程、评价结果 3 个维度, 运用 12 个观测指标对 16 份创新创业教育评价文本逐一进行内容分析, 结合问卷调查对当前创新创业教育评价文本的准确性、合理性、可行性和实用性进行元评估。结果显示, 所选的 16 份创新创业教育评价文本的元评估水平介于“一般”和“较差”之间, 合理性略好, 可行性和准确性次之, 实用性并不理想。基于以上分析, 从评价设计、评价过程、评价结果三方面提出改进建议。

[关键词] 创新创业教育评价; 元评估; 评价设计; 评价过程; 评价结果

[中图分类号] G640 **[文献标识码]** A **[文章编号]** 1674-893X(2022)06-0012-10

一、问题提出

近年来, 国家相继出台了《国务院办公厅关于深化高等学校创新创业教育改革的实施意见》《国务院关于推动创新创业高质量发展打造“双创”升级版的意见》等政策文件, 为高校创新创业教育的改革与发展指明了方向。为衡量教育成效, 高校纷纷构建了创新创业教育评价体系, 以求通过“以评促改”来提高创新创业教育质量。然而, 创新创业教育的评估体系能否发挥预期作用, 与采用的评估体系的质量有很大的关系, 因此, 为保证创新与创业教育质量, 有必要对创新创业教育采用的评估体系进行再评估。

元评估是人类的一种“自我参照”, 既可用于人类努力过程与努力成果的二度测量, 也可用于对评价本身的评价, 即“对评价的评价”^[2]。对创新创业教育评价文本进行元评估是推进创新创业教育评价规范发展的现实需要, 也是提升创新创业教育评价品质和效果的必然选择。

然而, 元评估并非可以确保评价体系的每个层面和环节合规、适切和最优, 但缺乏元评估制度架构的评估体系如在评价跑道上“盲跑”^[3], 就面临着跑偏的风险, 因此创新创业教育评价的发展呼唤元评估的审视。

为探究当前既有的创新创业教育评价体系的问题与不足, 本研究从元评估视角出发, 借助斯塔弗尔比姆的 CIPP 评价模式, 在对当前创新创业教育评价文本进行“评价设计—评价过程—评价结果”分析的基础上, 采用问卷调查的方法对创新创业教育评价进行元评估, 并基于评估数据, 对改进和完善创新创业教育评价提出政策建议。

二、研究设计与过程

(一) 元评估对象

当前关于创新创业教育评价的文本大多来自学者对创新创业教育评价体系的研究。因此, 在中国知网 (CNKI) 上以“创新创业教育评价”“创新创业教育效果评价”“创新创业教育评价

[收稿日期] 2022-03-09; **[修回日期]** 2022-09-23

[基金项目] 2020 年度国家社科基金教育学重点项目“促进教育治理能力提升的教育评价制度改革研究”(AFA200009)

[作者简介] 陈佳巍, 女, 河北保定人, 湖南大学教育科学研究院硕士生, 主要研究方向: 创新创业教育、工程教育认证, 联系邮箱: 1120918484@qq.com

指标”为篇名进行检索，剔除以高职院校为研究对象的文章，检索出相关文献共 113 篇。为探求评价文本在设计、实施、结果运用中的现状和存在的问题，本文在筛选文本时，仅保留构建了评价指标体系的相关文献，最终确定评价文本 16 份。

(二) 元评估内容

斯库瑞文(Scriven)提出元评估要重视评价原来的设计，检查数据分析的方法与过程^[4]。斯塔弗尔比姆(Stufflebeam)主张从评价主体是否得当、评价对象是否准确、评价过程是否科学合理、评价目的是否能达到等四个方面进行元评估^[5]。此外，美国评价协会将元评估流程划分为“规划

与协商、结构与设计、数据搜集与准备、资料分析与解释、沟通与公布、结果运用”^[6]等 6 大板块。以上研究表明，评估应考察设计、过程、结果等诸多方面。所以，本文欲从评价设计、评价过程、评价结果三个层面对评价文本进行元评估，具体评估观测点如图 1 所示。

本文从评估文本的目标定位、评价对象、评估指标结构、评价内容四个方面对评价设计进行研究。在评价过程层面，主要对评价的参与主体、评价技术运用、指标赋权情况、评价的过程监督情况进行探究，并从结果的表达、公布、对结果的反馈及运用维度进行评价结果的分析。

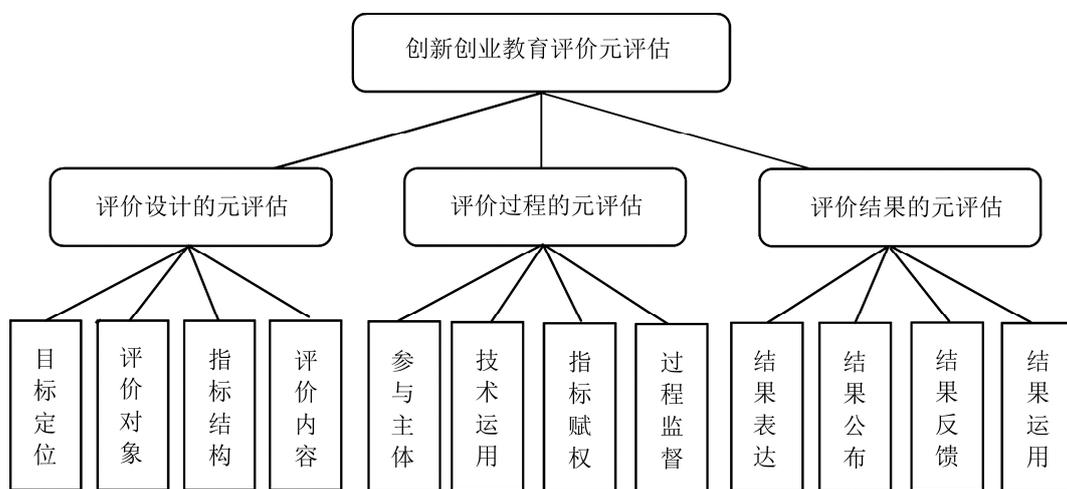


图 1 创新创业教育评价元评估的内容框架

近二十年来，美国及其它国家所进行的元评估，大多采用 1981 年美国教育评估标准联合委员会(JCSEE)开发的《方案评估标准》，主要包括效用性、可行性、适当性、准确性四个维度 30 项标准。经过检讨和修订，1994 年出版的第二版《方案评估标准》保留了效用性、可行性、适当性、准确性四个维度，但在内容上将原来的部分标准进行了合并，并增加了一些新的标准。亚伯勒、苏哈拉等人基于 JCSEE 的元评估标准，开发了包括效用性标准、可行性标准、适切性标准、精确性标准以及问责标准等五个维度在内的元评估标准^[7]。基于对以上元评估标准的梳理和思考，结合评价文本的具体内容，本研究从实用性、可行性、合理性、准确性四个维度对创新创业教育评价文本进行元评估，并对四

维属性的考察内容作出如下界定：实用性维度主要考察文本落地过程中对实践的指导程度，合理性维度主要考察文本中各评估指标的逻辑程度，可行性维度主要考察文本该项指标落地的困难程度，准确性维度主要考察该指标表述与现实考察的契合程度。

(三) 元评估程序与方法

本文在分析创新创业教育评价文本内容的基础上，制定元评估内容体系，并由此形成评估调查表，对文本各观测维度逐一进行评价。采用的研究方法具体如下。

1. 内容分析法

首先，对本文所选评估文本逐一进行内容拆解，分析各文本观测指标的类型，并进行标注。此外，为了调查表的制定与评估的开展，本研究

以句子为分析单位,采用“文本—一级指标—二级指标—(类别)—标准”的方式进行编码,如“1-1-4-(1)-1”表示对文本1评价设计中指标赋权(层次分析法)实用性的元评估,编码情况如表

1所示。将每份文本涉及的节点标记为“1”,文本未涉及则记为“-”。将编码结果在Excel中进行统计分析,进而得出评估文本结构的完整程度。

表1 文本1编码体系举例

文本	一级指标	一级指标编码	二级指标	二级指标编码	有无	三级指标	三级指标编码
1	评价设计	1-1	目标定位	1-1-1	1	实用性	1-1-1-1
						可行性	1-1-1-2
						合理性	1-1-1-3
						准确性	1-1-1-4

2. 问卷调查法

(1) 问卷编制

根据内容分析的四个维度12个观测指标进行通用问卷框架的制作,由此形成初始调查表。所有问题采用封闭式作答方式,采用李克特量表的五级评分模式,其中5分(最高分)代表“非常好”,4分代表“比较好”,3分代表“一般”,2分代表“比较差”,1分代表“非常差”。此外,本问卷设置了“0”分选项,仅当评价文本中不包含该项时才记为0分。初始调查表编制后,征集了5位创新创业教师的意见,对问卷的题项设置进行了优化,由此形成《创新创业教育评价文本元评估调查表》。

(2) 问卷调查实施

采用随机抽样法,向11名从事创新创业教育研究工作的专家发放调查表,邀请专家从各观测指标的实用性、可行性、合理性、准确性四个维度对16份文本进行五级评分。

本研究采用克伦巴赫 α 系数来检验调查表的信度。分析显示,元评估调查表的实用性维度为0.897,合理性维度为0.913,可行性维度为0.904,准确性维度为0.871,表明其信度较好。为检测调查表评定的信度情况,研究者随机抽取一份文本,在评定两周后再次邀请专家对该文本进行评定,测得斯皮尔曼等级相关系数 $r=0.87$,证明该调查问卷具有良好的稳定性。

信度分析之余,本研究对创新创业教育评价文本元评估调查表进行了效度分析。首先,《创

新创业教育评价文本元评估调查表》的观测指标经过了教育评价专家、创新创业教育一线教师的评判,保证了调查表的内容效度。此外,调查表四个维度之间的相关性以及维度与调查表总分间的相关情况如表2所示。

表2 调查表各维度之间与总分之间的相关情况

	实用性	可行性	合理性	准确性
实用性	1			
可行性	0.73**	1		
合理性	0.66**	0.57**	1	
准确性	0.59**	0.52**	0.67**	1
元评估总评	0.87**	0.83**	0.91**	0.88**

由表2可知,调查表各维度之间的相关系数介于0.52~0.73之间,基本上呈中度相关。4个维度与调查表之间的相关均高于维度间的相关,说明创新创业教育评价文本元评估调查表具有较好的结构效度。

(3) 数据统计方法

对调查表题项进行编码,将各编码得分情况在Excel中进行统计分析。三级指标得分由11位专家评分求算术平均而来(保留两位小数)。二级指标初始得分由其所对应的4个三级指标得分求算术平均得出(保留两位小数)。通过将二级指标初始得分与对应指标权重相乘求和,即可得出二级指标赋权得分(保留四位小数,以确保评分精确度)。综合各二级指标赋权得分,即可得出一

级指标初始得分。综合各一级指标赋权得分及其对应权重相乘求和，即可得出各文本最终的元评估得分。

创新创业教育评价元评估包括目标定位、参与主体、评价对象等 12 个观测指标，各观测指标对于文本的重要性程度不同，因此在处理调查数据前，本研究首先借助改进层次分析法 (Mufti-AHP)，对评价文本的 3 个一级指标及 12 个二级指标进行了赋权。经 11 位专家评测后，得出创新创业教育评价元评估体系的各指标权重，如表 3 所示。

表 3 创新创业教育评价元评估体系权重

	评价设计	评价过程	评价结果	数值指标
	0.335	0.333	0.332	与总排序
目标定位	0.247			0.082 7(7)
评价对象	0.236			0.079 1(11)
指标结构	0.234			0.078 4(12)
评价内容	0.283			0.094 8(1)
参与主体		0.253		0.084 2(2)
技术运用		0.251		0.083 6(6)
指标赋权		0.244		0.081 3(10)
过程监督		0.252		0.083 9(4)
结果的表达			0.253	0.084 0(3)
结果的公开			0.248	0.082 3(8)
结果的反馈			0.252	0.083 7(5)
结果的运用			0.247	0.082 0(9)

在分析内容的基础上，我们运用调查表进行样本评定，计算各指标维度的得分，根据得分所占的比重计算出得分率，并划分等级。当得分率大于等于 80% 时，元评估结果为“非常好”；得分率大于等于 60% 但小于 80% 时，元评估结果为“比较好”；得分率大于等于 40% 但小于 60% 时，元评估结果为“一般”；得分率大于等于 20% 但小于 40% 时，为“比较差”；得分率低于 20% 时，元评估结果为“非常差”。

三、研究结果与分析

(一) 内容分析的结果

内容分析的结果表明，评价主要采用“指标量化评语描述”方式，但部分文本的一些观测信息有缺失。由于本研究是基于评价指标体系进行

元评估探讨的，所以在 16 份评价文本中，评价的“目标定位”“评价对象”“指标结构”“评价内容”“参与主体”几乎都有所体现。从元评估的二级指标来看，文本虽然较多地探讨了“评价内容”（如指标体系构建）、“评价标准”等设计层面以及“技术运用”“指标赋权”等过程层面的内容，但是对评价过程层面的“过程监督”、评价结果层面的“公开”“反馈”“运用”等内容规定较少，且多在讨论中以“建议”的形式呈现，文本的完整性有待提高。从一级指标来看，文本较多地关注了评价设计，其次是评价过程，而对评价结果的关注较少。具体情况如表 4 所示。

表 4 各文本观测指标包含情况统计表

二级指标	覆盖率(%)	备注
目标定位	87.50	
评价对象	100.00	
指标结构	100.00	
评价内容	100.00	指标详细说明覆盖率 56.25%
参与主体	100.00	
技术运用	62.50	
指标赋权	75.00	
过程监督	43.75	多见于文本的政策建议中
结果的表达	56.25	
结果的公开	31.25	多见于文本的政策建议中
结果的反馈	31.25	多见于文本的政策建议中
结果的运用	37.50	多见于评价目的中

从评价的目标定位来看，超过九成的文本在评价目的文字表述上体现了“提高创新创业教育质量，培养创新型高素质人才”的培养目的，如为完善大学创新创业教育提供决策^[8]，为提高创新创业教育质量评估水平提供新的思路和方法^[9]，从而提升高校创业教育水平，以“评”促发展^[10]，使高校创业教育实现价值最大化^[11]等。但从评价过程和结果使用来看，能够将“以评促改”贯穿始终的文本不及三分之一，不少评估意在呈现结果。

从评价对象来看，虽然多数文本清晰地指向高校创新创业教育质量、绩效或现状，但是约有四分之三的文本并未对高校的级别、类型做细致的划分，多数文本尝试建立通用的评估框架。从评价指标所涉及的参与主体来看，评价工作多由

高校做顶层设计,参与更多的仍然是一线教师,仅将用人单位、毕业生、校友的“满意度”作为创新创业教育评价的信息补充。

从指标的筛选与构建来看,在所评估的16份文本中,有68.75%的评价文本采用因素分析法,43.75%的评价文本采用调查法,31.25%的文本使用评价模型(有的文本可能采用两种或两种以上的研究方法)。因素分析法主要依据现有研究成果,将创新创业教育目标进一步分解和细化为指标。该方法多凭借研究者的经验来分析,有较强的主观性,如果评审人员选择不当将对评价结果的信度产生负面影响。当构造判断矩阵的指标数目较多(超过9个)时,容易造成评审专家的判断混乱,一致性这一要求也将难以满足。基于以上缺陷,部分文本结合调查法,通过向有关专家进行调查与咨询、向学生和教师开展调查和访谈,由此完成指标筛选和确定。如在充分借鉴国外相关研究成果的基础上,结合我国创新创业教育的发展现状构建评价指标体系,通过向同行、有关专家进行问卷调查^[12],完成指标的初选、优选和最终确定;或基于扎根理论,抽取6省的政府、高校、企业相关人员,围绕高等学校创新创业教育评价究竟“评什么、如何评、当前评价困境及如何解决”等问题进行半结构访谈、焦点小组访谈、小型座谈会,将收集资料进行编码分析,得以构建创新创业教育评价结构模型;也有文本在CIPP评价模型(创业环境基础能力、资源配置能力、过程行动能力和成果绩效能力^[13])、DPSIR评估模型(驱动力、压力、状态、影响、响应^[14])、解释结构模型(生源质量、创新创业师资队伍、创新创业教育质量、学员成长过程、应届毕业学员创新创业质量、毕业学员中长期创新创业发展质量、学员创新成果与创业评价^[15])、柯氏评估模型等框架之上构建了评价指标体系。评价体系的结构以两层和三层居多,建立三层指标体系的文本占比56.25%,不到六成的文本在指标体系中列出了评估所需的资料和方法等操作性说明。

从评价的技术运用来看,评价信息主要通过规范计分和人员评定两种评价技术进行处理。第一种评价技术为规范计分,12.5%的文本采用直

接计分法,6.25%的文本采用按等给分,25%的文本采用分等划分。直接计分式关注显性指标和非显性指标特点的不同,将显性指标的落实情况(如个数、比率)记作该指标的得分,如“核心课程开出率”为32%,指标对应权重为0.0319,“企业家访问演讲数”为22场,指标对应权重为0.0183,则综合评价为 $32*0.0319+22*0.0183$ ^[16]。但因各指标有不同单位和数值范围,所以仅将各指标的评价结果进行横向比较缺乏科学性,且从各指标的评价分数中较难得出创新创业教育的质量。也有文本设定每个指标的最高分数,采取扣完为止的形式确定各指标最终得分。如“学术会议评分以参加省级学术会议人次数计数算分,5人次以上得5分,少1次扣1分,扣完为止”^[17]。按等给分式体现为对各指标划分等级,为每个等级设定固定分数,如“课程开设数量”为3门以下为不合格,计10分;4~5门为基本合格,计30分;6~7门为合格,计50分;8~9门为良好,计70分;10门以上为优秀,计90分。分等划分式是对各指标划定等级并列明分数区间,根据各指标内涵进行计算,以计算结果位列该指标等级来给定指标得分。如用“创业社团”数量反映创业社团的社会影响力,若社团数量多于4个,则这一指标评价结果为“优秀”,给予90~100分之间的分数,若社团数量少于3个,则这一指标评价结果为“一般”,给予70分以下的分数,否则评价结果为“良好”,给予70~80分之间的分数。第二种评价技术为人员评定,在所评估的16份文本中,有6.25%的文本采用直接评分的方法,12.5%的文本采用模糊评级法。直接评分是专家根据每个指标的现实情况,给予各指标0~100分的评分。模糊评级是采取模糊评价法,向评审人员(如高校高校主管创新创业教育工作的领导、学生、教师等)发放调查问卷,基于其自身的经验或感受,选出分评价指标实施情况的程度式评语,由此构建单因素的模糊综合评价矩阵,再综合专家估测所得的各指标权重^[18],确定各分指标的隶属度,在复合运算基础上,得出综合评价等级。

从评价指标的赋权来看,有四分之三的文本

进行了指标权重的确定,层次分析法、改进层次分析法、群组 G1 法等是确定指标权重的常用方法。层次分析法是通过把复杂问题中的各种因素划分为相互联系的有序层次(即目标层、准则层、方案层),利用专家或创新创业教育利益群体的经验和判断能力,对各指标进行两两比较。该方法将定性指标和定量指标进行统一比较,规避了不可量化指标其权重难以确定的问题^[19]。部分文本将群组序关系分析法(群组 G1 法)引入创新创业教育评价,通过对指标进行指标类型一致化和指标无量纲化处理,选择四位(通常为两位以上)不同领域的专家进行指标排序^[20],根据专家序关系一致或不一致的情况,采取不同方法计算权重。此外,也有文本采用熵权和逼近理想解的排序方法(TOPSIS)相结合、专家估测法、德尔菲法、问卷调查等方法进行指标权重的确定。

从评价结果来看,当前创新创业教育评价主要通过评分和评级两种方式呈现。绝大多数文本对评价结果的公开和反馈并未作出明确规定,仅有 31.25%的文本以建议的形式提出公开的评价数据与评价结果,以及加强过程监督评价。由于所选样本更多地关注评价指标体系的设计和评价结果的量化计算,并非为完整的评价文本,在评价结果的使用、反馈等环节还存在明显的不足。

从整体上来看,部分创新创业教育评价文本呈现出以下特点:明确指出了评估目的,依托理论基础或实际调研构建量化评价指标体系,通过指标赋权和量化评分或评级确定评价结果。由于政府是高校开展创新创业教育资源的主要供给者,得到政府和社会的认可依然是高校追求的首要目标^[21],因此不难发现,现有的评价指标体系较多地基于高校视角来呈现高校“做了什么”“哪里做得好”以及学生“学了什么”等方面的可量化指标,如对师资队伍建设的投入、学校的制度建设、课程建设、教学管理、教学资源的提供和经费保障等方面,而对“学生学得如何”等教育成效指标的关注则不足,且缺乏对创新创业教育反馈情况的评价及“后设评估”,多数文本对指标进行评价的操作化阐释不明。此外,评价重“短

期”轻“长期”,评价内容多为短期指标。创新创业教育具有时滞效应,“企业创业者的平均年龄为 33~35 岁,大约在完成大学教育后的十年左右进行创业”^[22],毕业生当年的创业率和创业意向只能测评创新创业教育的部分短期效应,而对创新创业教育的长期效果及滞后性反映不足。创业数、对社会经济发展的贡献程度等指标往往在创业教育结束较长时间后才能够进行测评^[23],测评的难度较大致使现有文本忽略或避开了这类长期指标。多数文本所构建的通用评估框架在评价内容设置上较少考虑大学间的差异性乃至专业的异质性,在内容表达上对区域特色、大学的发展定位突出不够,“以评促改”的效果有待进一步审视。尽管近年来,多元主体参与创新创业教育评价的意识相较之前有所增强,但用人单位、毕业生、校友等评价主体仅作为评价的信息补充者,以“满意度”这一单一指标作高校主体评价的少量补充,评价主体的“主体性”体现仍然不足。此外,从文本的完整性来看,文本在评价的过程监督、评价结果的使用和反馈等方面体现不足。

(二) 四维量化调查结果

调查数据表明,16 份文本的平均得分率在 43.72%±15.11%上下浮动,评价等级介于“一般”和“比较差”之间,评价文本整体水平偏低。评价的合理性略好,可行性和准确性次之,实用性并不理想,四者之间的差异并不显著。综合观测指标的四维得分与对应权重,可以得出 16 份创新创业教育评价文本的最终元评估得分为 2.455 5,得分率为 49.11%,元评估等级为“一般”。

从元评估的一级观测指标来看,“评价过程”的元评估结果评级为“比较好”,得分率略大于 60%，“评价设计”“评价结果”的元评估结果评级为“一般”,且位于该等级的中下水平。细化到元评估的二级观测指标可以发现,5/6 的观测点等级为“一般”及以下。尽管“指标赋权”“目标定位”等观测指标的元评估结果比较理想,但尚未达到“非常好”等级。“技术运用”“结果的表达”“参与主体”等指标的整体得分也相对较高,得分率位于“一般”评级的得分前

列。“结果的公开”“结果的反馈”“过程监督”在评价工作中亟待完善与改进，当前尚处于“比较差”的等级。

从元评估二级观测指标的四维特性来看，评价设计的合理性最好，可行性和实用性次之，准确性最差，但三者之间得分差别较小；评价过程

的合理性最好，可行性和实用性次之，准确性最差，合理性与准确性之间有较大差别；评价结果的合理性最好，实用性和可行性次之，准确性最差，评价结果合理性的得分率几乎为准确性得分率的两倍。12个二级观测指标元评估得分情况如表5所示。

表5 观测指标元评估得分情况统计表

一级指标	赋权分	二级指标	实用性	可行性	合理性	准确性	均分	赋权分
评价设计	0.808 3	目标定位	3.07	3.03	3.21	2.94	3.062 5	0.756 4
			61.40%	60.60%	64.20%	58.80%	61.25%	
		评价对象	1.77	2.27	3.15	1.64	2.207 5	0.521 0
			35.40%	45.40%	63.00%	32.80%	44.15%	
		指标结构	2.41	2.07	2.41	2.01	2.225 0	0.520 7
48.20%	41.40%		48.20%	40.20%	44.50%			
评价内容	1.92	1.94	3.07	1.76	2.172 5	0.614 8		
	38.40%	38.80%	61.40%	35.20%	43.45%			
综合	2.283 3	2.317 5	2.969 0	2.081 6	2.412 9	2.412 9		
	45.67%	46.35%	59.38%	41.63%	48.26%		48.26%	
评价过程	0.914 2	参与主体	—	3.01	2.81	1.47	2.430 0	0.614 8
			—	60.20%	56.20%	29.40%	48.60%	
		技术运用	3.22	2.48	3.29	2.63	2.905 0	0.729 2
			64.40%	49.60%	65.80%	52.60%	58.10%	
		指标赋权	3.83	3.71	3.75	3.53	3.705 0	0.904 0
76.60%	74.20%		75.00%	70.60%	74.10%			
过程监督	1.41	2.27	2.24	—	1.973 3	0.497 3		
	28.20%	45.40%	44.80%	—	39.47%			
综合	2.098 1	2.861 3	3.016 2	1.893 4	2.745 2	2.745 2		
	41.96%	57.23%	60.32%	37.87%	54.90%		54.90%	
评价结果	0.733 0	结果的表达	2.34	—	3.21	2.26	2.603 3	0.658 6
			46.80%	—	64.20%	45.20%	52.07%	
		结果的公开	1.76	1.44	2.71	1.42	1.832 5	0.454 5
			35.20%	28.80%	54.20%	28.40%	36.65%	
		结果的反馈	1.54	2.24	2.47	1.61	1.965 0	0.495 2
30.80%	44.80%		49.4%	32.2%	39.30%			
结果的运用	1.97	2.67	3.06	2.01	2.427 5	0.599 6		
	39.40%	53.40%	61.20%	40.20%	48.55%			
综合	1.903 2	1.581 1	2.862 5	1.474 0	2.207 9	2.207 9		
	38.06%	31.62%	57.25%	29.48%	44.16%		44.16%	
最终得分	2.455 5	2.094 8	2.253 3	2.949 2	1.816 3	2.455 3	2.455 3	
	49.11%	41.90%	45.07%	58.98%	36.33%	49.11%		49.11%

四、未来展望

评价文本的制作要力求严谨、科学，内容和形式应合乎规范。可以通过制作与文本配套的评估手册、评估指南等解读性文本来提高评价文本的可行性和准确性。要加强创新创业教育质量评价的服务性，提高质量评估的科学性。具体展望如下。

第一，创新创业教育评价设计：明确理念，分类评价，深挖内容，定量结合。明确人才培养理念和评价目标，加强创新创业教育分类评价与指导。创新创业教育的最终目标是在现有的教育资源下，为地方乃至国家最大限度地培养创新型建设人才，而评价的根本目的是为了检验和反馈这一目标的实现情况。高校应面向社会征询具体的人才培养办法，并根据本校的办学定位和所在地域的特色优势，做好顶层设计，明确创新创业教育工作的短期、中期及长期规划，使创新创业教育评价体系更有针对性和目的性。此外，应针对不同的学生群体，建立本科生、研究生等分类评估体系；基于专业划分类型的不同，建立工科专业、医科专业、财经专业以及师范专业等分类评估体系；充分考虑院校类别的差异，建立“双一流”院校、地方院校、民办高校等分类评估体系。由于创新创业教育的复杂性，在构建评估体系时不应仅着眼于某一类标准，应在充分考虑被评高校创新创业教育的级别、特点、目标、内容、结构的基础上，综合学生的家庭背景、个性特征^[24]以及高校所在地区的经济特色等多重因素，构建创新创业教育分类矩阵，以目标为导向，作出合理评估。

深挖评价内容，定量和定性评价相结合。在制定评价文本时，必须充分考虑被评高校创新创业教育的特点，在明晰创新创业教育的作用机制和有效的边界条件的基础之上选择与确定评价指标^[25]，避免指标体系同质化、概念化。评价指标体系应充分反映行业特点。要创新评价形式，摆脱“数据见真谛”的量化指标陷阱，评价维度不仅要包含可量化的客观要素，也应扩充不可用数据描述等无形要素，如校园文化、办学理念、地方定位、历史传承等。要根据各要素特点，

制定不同的评价标准和方法。要建立系统化、全方位、动态可持续的创新创业教育评价体系，深化纵向评价，将第一年数据作为基准数据，通过对比近年来各指标的纵向评价数据，找寻调整教育投入的方向。要注重对创新创业教育的反馈情况进行评价与监督，将“反馈机制”纳入评价指标体系，如从“是否为学生提供便利有效的创新创业教育学习需求反馈渠道”“是否对学生的创新创业教育学习情况进行跟踪”等方面进行评价^[26]。

第二，创新创业教育评价过程：多元参与，优化方法，规范过程，加强监督。以网络为媒介，实现多主体参与的综合评价。评价体系作为一种信息反馈体系，应在大数据时代下实现与时俱进。高校应充分挖掘数据信息，借助网络平台，收集多个评价主体的信息，关注创新创业教育对人才的全过程培养。加大社会参与的深度和广度，让社会参与监督和评价创新与创业教育的质量，将创新创业教育效果适应并满足社会需要的程度作为质量评价标准。建立政府、高校、第三方评价机构三位一体的多元主体评价方式，充分发挥第三方评价机构的专业性和灵活性，在高校自评之余科学地反映存在的问题与不足^[27]。

优化信息收集与处理方法，在“传统”与“现代”中碰撞出火花。随着科学技术的发展，大数据逐渐应用于教学过程中，教育信息化已经成为推动教育高质量可持续发展的重要手段。获得完整、真实的数据和证据是教育评估有效实施的保障，且创新创业教育涉及校内外多个评价主体，因此在传统评价方法之外，还要充分借助大数据、人工智能等“现代”手段，全方位、多元化地采集数据，分析信息。

制定详细合理的评价标准，规范评价的各个环节。完善创新创业教育评价操作手册，在评价设计、评价过程和评价结果输出等方面加强程序规范，用通俗、严谨的文字予以明确。加强对评估人员的思想与技能培训，增强评估人员的信效度意识，提高评估人员对科学评估的重视程度。在对评估人员进行大数据相关知识和理论的学习中提升其大数据意识，使其熟练地运用信息技

技术手段,将大数据技术熟练应用于创新创业教育与评价工作中,实现创新创业教育评价工作的减负增效。此外,为保证评估人员评分的一致性,可在正式评估前,对评估人员进行预评估。

加强评价过程的监督管理,责任到人。建立评价过程的监督管理制度,对评价工作的开展标准予以明确,依据评价人员评价工作的落实情况进行合理奖惩。信息的收集和处理资料要注明来源、评价方法和评审人员,提交评价结果时应一并递交原始资料,以便同行或管理人员进行抽检复核。

第三,创新创业教育评价结果:公开管理,规范报告,加强反馈,落实改进。公开管理评价结果,规范评价结果报告。做好评价信息的管理工作,对所采集的数据、证据进行归档留存,将采集的过程和评价结果予以公开说明,力求做到科学和透明。在结果的输出上,规范评价结果报告的内容及格式,以精确、简明的语言呈现各指标评价结果,让学生、家长、用人单位、企业等利益主体了解和掌握评价结果,根据评价信息作进一步决策。

有效反馈评价结果,落实评价结果运用。新时代教育评价的最终目的指向人才培养质量和学校办学水平的提高,而实现该目标的关键在于科学评价之下对结果的有效反馈和持续改进。高校应不断健全与完善创新创业评估机制,坚持对评估的评估,以发展的眼光分析评估设计、评估实施过程以及评价结果的反馈与运用,分析评估工作的实用性、可行性、合理性、准确性,提高评估的价值,使评估效益更大化。此外,因材施教同样适用于创新创业教育。依托校园网搭建创新创业大数据共享平台,链接创新创业教育各参与主体,设置评价反馈板块,通过评价结果的反馈,引导学生反思和总结自身的学习情况,达到改进自身学习的目的。将评价权限合理分配给评价体系中相关的各大评价主体,定期收集、整理和分析反馈信息。网络后台根据学生关注的板块、自媒体推送的信息、互动人群和话题等综合分析学生的学习兴趣和创新创业意向,对学生进行分类管理,为其提供精准化学习

和实践服务。

第四,本研究局限及进一步研究方向。本研究在对评价文本进行内容分析的基础上,从元评估视角出发,选取专家作进一步问卷调查,分析得出创新创业教育评价文本的现状,这一创新对创新创业教育评价的建设与完善开拓了新的思路。然而,科学合理的创新创业教育评价需要高校、教师、学生、家庭、社会等多元评价主体共同参与,仅选取专家进行评价的再评估有较大的片面性。之后的研究将在扩充评价主体的基础上,综合各评价主体评价结果的权重,进行更全面深入的分析。此外,如何增加评价工作透明度,保证参评信息的可获得性和准确性,形成常态化的内外考评机制是今后研究的重点。

参考文献:

- [1] 梁正瀚. 新时代创新创业教育评价体系构建研究[J]. 中学政治教学参考, 2021(23): 106-107.
- [2] SCRIVEN M. Types of evaluation and types of evaluator[J]. Evaluation Practice, 1996, 17(2): 151-161.
- [3] 高文杰. 元评估: 我国高职教育评估亟待引入的制度架构——基于新制度经济学的视角[J]. 职教论坛, 2016(7): 52-57.
- [4] SCRIVEN M. Meta—evaluation revisited[J]. Journal of multi disciplinary evaluation, 2009(11): 3-8.
- [5] STUFFLEBEAM D L. The meta-evaluation[J]. American Journal of Evaluation, 2001(2): 66-69.
- [6] WINGATE L A. The program evaluation standards applied for meta - evaluation purposes: investigating interrater reliability and implications for use[J]. Proquest Llc, 2009: 181.
- [7] YARBROUGH D B, SHULHA L M, HOPSON R K, et al. The program evaluation standards: A guide for evaluators and evaluation users[EB/OL]. (2013-10-29) [2020-12-28]. <http://www.eval.org/p/cm/ld/fid=103>.
- [8] 王宇鹏. 全过程视角下基于 AHP 的大学生创业教育评价研究[J]. 渤海大学学报(自然科学版), 2014, 35(4): 398-402.
- [9] 黄金鹏, 刘江雁. 创业教育质量模糊评估体系设计研究[J]. 辽东学院学报(社会科学版), 2015, 17(3): 131-135.
- [10] 邱丽华. 高校创业教育评价体系构建及应用研究[J].

- 学校党建与思想教育, 2015(13): 81-83.
- [11] 黄海珍. 基于平衡计分卡的创业教育绩效评价[J]. 岳阳职业技术学院学报, 2015, 30(6): 30-33.
- [12] 刘海滨, 杨颖秀, 陈雷. 基于 AHP 的大学生就业创业教育评价指标体系构建[J]. 东北师大学报(哲学社会科学版), 2012(6): 227-232.
- [13] 葛莉, 刘则渊. 基于 CIPP 的高校创业教育能力评价指标体系研究[J]. 东北大学学报(社会科学版), 2014, 16(4): 377-382.
- [14] 武冠蓉. 基于 DPSIR 模型的高校创业教育质量评价体系建设研究——以某应用型本科院校为例[J]. 山西高等学校社会科学学报, 2019, 31(1): 77-80.
- [15] 王兴立. 大学生创新创业教育质量评价的矛盾困境、体系优化及创新举措[J]. 教育与职业, 2018(4): 68-72.
- [16] 秦敬民. 基于 QFD 的高校创业教育质量评价研究[D]. 天津: 天津大学, 2010.
- [17] 张丹译. 高校创新创业教育绩效评价研究——基于湖北 47 所本科院校的实证调查[D]. 武汉: 武汉科技大学, 2019.
- [18] 徐英, 白华. 高校创新创业教育绩效评价研究[J]. 创新与创业教育, 2014, 5(2): 29-33.
- [19] 姚倩, 韦颖. 高校创业教育评价实施路径探析——基于 66 篇文献的定量研究[J]. 创新与创业教育, 2019, 10(4): 84-88.
- [20] 李旭辉, 胡笑梅, 汪鑫. 高校创新创业教育效果评价体系研究——基于群组 G1 法的分析[J]. 教育发展研究, 2016, 36(21): 29-36.
- [21] 白逸仙. 走向“以学生为中心”的评估模式——以中国《本科教学质量报告》与美国 NSSE 为比较对象[J]. 中国高教研究, 2014(11): 64-68.
- [22] 沈超红, 陈彪, 陈洪帅. 创业教育“时滞效应”与创业教育效果评价分析[J]. 创新与创业教育, 2010, 1(4): 3-7.
- [23] 梅伟惠. 高校创业教育评价的类型与影响因素[J]. 教育发展研究, 2011, 31(3): 45-49.
- [24] 陶丹, 陈德慧. 中国高校创业教育质量评价指标体系研究[J]. 科技管理研究, 2010, 30(5): 84-86.
- [25] 王占仁, 刘志, 刘海滨, 等. 创新创业教育评价的现状、问题与趋势[J]. 思想理论教育, 2016(8): 89-94, 103.
- [26] 杨月涵. 学生视角下高校创新创业教育评价指标体系研究[D]. 天津: 天津理工大学, 2019.
- [27] 田夏. 高校创业教育评价体系的类型与构建[J]. 教育探索, 2016(4): 82-85.

The present situation of evaluation of innovation and entrepreneurship education in colleges and universities from the perspective of meta-evaluation

CHEN Jia-wei

(Education Science Research Institute, Hunan University, Changsha 410082, China)

Abstract: The improvement of the quality of innovation and entrepreneurship education should be based on scientific evaluation. With the help of CIPP evaluation model, from the three dimensions of evaluation design, evaluation process, and evaluation results, applying 12 observation indicators to analyze the extents of the 16 copies of innovation entrepreneurship education evaluation scheme respectively, and combined with the questionnaire survey, this paper makes a meta-evaluation of the accuracy, rationality, feasibility and practicability of the current innovation entrepreneurship education evaluation scheme. The results show that the meta-evaluation level of the 16 selected innovation and entrepreneurship education evaluation programs is between “general” and “poor”, the rationality is slightly better, the feasibility and accuracy are second, and the practicability is not ideal. Based on the above analysis, suggestions for improvement are put forward from three aspects of evaluation design, evaluation process and evaluation results.

Key Words: innovation and entrepreneurship education evaluation; meta-evaluation; evaluation design; evaluation process; evaluation results

[编辑: 游玉佩]