

第一章

职业教育的学业评价

一、学业评价方法：一个被忽视的重要领域

科学的教育评价对职业教育的健康发展具有重要的意义。2020年10月，中共中央、国务院印发了《深化新时代教育评价改革总体方案》（以下简称《改革方案》），提出建立富有时代特征、彰显中国特色、体现世界水平的教育评价体系的奋斗目标。针对学业评价，该《改革方案》提出完善“学生学业要求”“过程性考核与结果性考核有机结合的学业考评制度”，以及“高等职业教育‘文化素质+职业技能’考试招生办法”等多方面的任务，这对职业院校学业评价工作提出了更高的要求。

学业评价是指以教育教学目标为依据，运用恰当、有效的工具，系统收集学生在学习过程中认知行为变化的信息和证据，根据一定标准对学生的知识、技能和能力水平做出客观衡量和科学判定的过程。合理的学业评价要体现党和国家的教育方针，尊重和体现学生的个体差异，并激发学生努力实现自身的价值。普通教育学业评价相对简单，通常根据课程标准规定的学习目标，采用相应测量工具进行，如标准参照性测验和成就测验等^①。职业教育既要迎合劳动力市场的功利性要求，又要实现促进学生生涯发展的教育性目的，其学业评价更为复杂和多样，除课程考试外，还有技能竞（大）赛、对口升学考试和职业能力测评等多种形式。与普通教育相比，职业教育学业评价结果的影响也更大，它直接影响着学生的升学和就业，影响着企业员工招聘及职业院校的发展。目前职业教育的学业评价在方法层面尚存在不少问题，主要表现在以下方面。

^① 钟启泉，汪霞，王文静. 课程与教学论 [M]. 上海：华东师范大学出版社，2008：293.



（一）传统学历教育的课程评价脱离实践

院校内的科目考试缺乏可比性，一些地区通过专业测评建立教学质量监控系统，也遇到方法的科学性和经费可及性等困难，即便是发达地区也是如此。例如，广州教育局曾在所属职业学校针对开设最多的 10 门专业基础课和 2 门专业核心技能课进行检测，但检测科目也仅能覆盖全部课程的 18%，难以全面反映学生的综合学业成绩^①，后未能坚持下来。

（二）X 技能等级证书缺乏规范的技能考核方法

对已发布的 X 职业技能等级标准进行文本分析发现，目前技能评价组织对技能评价方法的科学性尚不敏感，采用的评价方法有些强调操作技能考试（如特殊焊接技术），有些只提到了知识考核（如失智老年人照护），有些尽管关注理论和实践相结合但依然采用笔试，实操考试也有笔试、网络考试等多种方式，一些证书甚至没有提出明确的考核评价方法。

（三）职教高考科学性不足，“指挥棒”作用有待优化

在各地开展的职教高考中，技能考试多针对特定的操作活动，关注技能的熟练度和可测量性，关注该项技能对整体职业能力的甄别和促进作用以及未来工作中的普适性，则少有系统性考量，而且标准含糊，打分随意^②；专业课考试命题缺乏行业和考试专家参与，难以保证考试效度和区分度^③。由于缺乏有效的技能测试，很多职教高考的内容主体仍然是文化课和专业理论，这既无法反映行业企业的用人需求，也背离了职业教育宗旨，造成职业院校对专业技能教学的不重视。

（四）技能大赛存在评价效度问题

调查发现，有的技能大赛获奖毕业生职业发展后劲不足，其综合素质与没有参加竞

① 柳洁，陈泽宇．SOLO 分类理论在职业教育学业评价领域的应用剖析 [J]．中国职业技术教育，2018(17)：5-10.

② 陈江．高职院校分类考试招生改革样态：问题与策略 [J]．高教探索，2019(2)：97-102.

③ 王乐，林祝亮．浙江省技能高考的问题及对策研究 [J]．河南科技学院学报，2018(8)：35-39.

赛的其他同学相比甚至处于劣势^①。这说明,如果评价方式有缺陷,技能大赛也无法实现提高人才培养质量的初衷。个别省份开展的技能抽查的科学性问题就更多了。

鉴于职业教育学业评价的难度,很多地区把文化课统考作为职业院校(特别是中职)学业评价的主要方式,这有可能造成职业教育“去职业化”趋向,在此美国提供了前车之鉴。2002年,当时的布什政府推出《不让一个孩子掉队法案》,试图解决美国中小学教育质量低下的问题。该法案的实施,虽然提高了学生的阅读和数学成绩,尤其是处境不利的学生,但是极大地削弱了职业课程,甚至很多学校放弃了非考试科目课程,其中就包括很多职业教育课程^②,导致美国中等职业教育事实上的“消失”。目前,强调文化课统考的《不让一个孩子掉队法案》已经被《每一个学生成功法案》取代,后者提出在基础教育阶段重新发展多样化的课程(包括生涯与技术教育),这为美国重振中等职业教育提供了可能。

综上所述,在职业教育质量保障体系建设过程中,开展科学、规范且实施便捷的学业评价的难度长期被低估,这也凸显了《改革方案》的及时性和重要性。职业教育学业评价必须反映技术技能人才的发展性目标和职业规范的双重要求,这对评价方法提出了巨大挑战。本书将探讨具有职业教育特色的学业评价的方法性特征,以期引发进一步讨论并为具体评价方案设计提供理论基础。

二、职业技能考试和职业能力测评

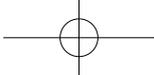
职业院校学生学业评价离不开对学生职业能力或技能水平的评价,其基本方式是职业技能考试和职业能力测评。

(一) 职业技能考试

不同国家有不同的职业技能考试制度,但大多会采用“知识+实践技能”的考核方式。

^① 苏敏. 中职汽车运用与维修技能大赛获奖毕业生追踪调研[J]. 中国职业技术教育, 2020(11): 87-92.

^② 高原. 美国当代标准化测试的命运与教育权利的转移[J]. 课程·教材·教法, 2016(9): 121-127.



职业技能评价方法研究

我国的职业技能等级（资格）鉴定考试一般也分为理论和实践考核两部分，采用纸笔测试、现场考核、典型作业、模拟操作等形式。职业技能技巧主要通过专家观察评分法确定；对职业规范和工作道德等隐性范畴的内容，则通过对工作过程和工作结果的观察进行考评。国外研究发现，目前职业技能评价中普遍采用的标准化考试的职业效度并不高，因为企业真实的生产服务活动不是简单的操作或线性过程，不能简单按照“对—错”标准或“投入—产出”关系来衡量^①。

技能大赛是一种社会关注度高、高利害性且特殊的技能考试，不同大赛在赛项设置、评分标准等方面有所不同。职业院校技能大赛常围绕教学标准及生产技术设置赛项，强调工艺难度，这与世界技能大赛有较大区别；世界技能大赛被认为对专业技术水准和职业素质能够进行更为全面的检验^②，实施成本高昂，但是如果其试题转化应用得好，也可以为职业教育的高质量发展提供有效的借鉴。

当今社会，技术发展对传统的技能考试产生了重要影响。人们发现，除了特别简单的动手操作行为（如点钞或切土豆丝等），采用主观性较强的观察法对复杂技能评价的结果的信度很低，因此在发达国家技能考试中出现的一个共同趋势是采用“表现性评价”，即考生“在特定情境中运用获得的知识完成某项任务，以考查其知识技能掌握程度，或问题解决、交流合作和批判性思考等多种复杂能力的发展状况”^③。此外，技能的含义也得到了很大拓展，人们更加强调应用知识完成任务和解决问题的能力。《国家职业教育改革实施方案》（简称“职教20条”）明确指出职业技能等级证书要“反映职业活动和个人职业生涯发展所需要的综合能力”，这说明在“1+X”技能证书制度设计中，“技能”和“能力”已经几乎成为同义词，职业技能正在发展成为更强调认知和综合性的职业能力。此外，传统的技能考试费时费力，特别是在开展大规模的技能测试时，要想达到较高的信度和效度要求，测试费用和人员投入会高到令人无法接受的地步（如奥运会体操等项目要求多名裁判同时评分）。因此，很多（大规模）职业技能考试需要向成本较低

① Young M. National Qualifications Frameworks as a Global Phenomenon[J]. Journal of Education and Work, 2003, 16(3): 223-237.

② 刘东菊. 世界技能大赛对我国职业院校人才培养启示[J]. 中国职业技术教育, 2012(36): 48-52.

③ 姐媛媛. 美国真实性学生评价及其启示[J]. 教育发展研究, 2007(6): 62-66.

的职业能力测评方向发展（详见第二章）。

（二）职业能力测评

职业能力测评旨在评价应用知识技能解决实际工作中的问题的能力，它与职业技能考试的区别在于：考试针对教学（或职业）标准，考察对学习内容的掌握程度，一般有对错之分，反映教学标准的效度；能力测评评价与特定工作的认知特征和认知水平相关，考察实际工作能力，只有能力水平高低之分而没有标准答案，反映职业要求的效度。

职业能力测评的基本功能是：将能力发展的实际水平及时反馈给教师和学生，帮助其处理学习过程中发生的问题并进行过程优化，是鼓励先进、鞭策后进和调动学习积极性的激励手段。当今社会对技术技能人才的全面发展和综合性要求越来越高，技术技能工作者必须有能力（参与）设计工作计划，按照经济性、创新性和环境可持续发展等综合性要求对工作的过程和成果进行评价。科学的能力测评要用整体化观念看待职业能力，不仅关注工作任务的完成情况，而且关心学生是否能够对工作过程和工作成果进行反思和改进，科学的能力测评应加强对诸如技术敏感性和创新性等高层次实践能力的的评价，促进学生的“价值理性”和“事实性评价能力”的发展。

从宏观管理角度看，如果对职业院校学生进行的职业能力测评能够达到大规模质量监控的信度和效度要求，就可以借此进行学校之间甚至区域之间的比较，并对不同学校人才培养过程的优劣进行分析和解释，从而为政府的科学决策提供实证基础。在此，需要研究和开发基于先进学习理论（如建构主义和情境学习理论等）的、技能考试与职业能力测评相互补充的、反映综合能力评价要求的、具有高预测效度的评价模型和评价实施方案，并为职业院校和技能评价组织设计和实施高质量的评价提供实操性的工具和实践经验。

三、学业评价标准

个体的学业成就是学习过程的产出或结果，学业评价旨在对学习者的学业成就水平的高低（学习绩效）进行判定和分级，处于某一级别的人必须满足相关标准的质量要求。



职业技能评价方法研究

标准是判断质量优劣和满意程度的门槛性要求，可以是一种标杆、价值取向或一项最佳成就记录，如精确性、准确性或与某一样板（人或事）的相符程度。在教学实践中，很多教师往往寄希望于制定绝对“科学合理”的标准，这个目标可能永远也无法实现。莱茵伯格（F. Rheinberg）将学业成就评价标准分为三类，即事实性标准、社会性标准和个性化标准^①，在此基础上可以界定具体评价指标，其中只有事实性标准属于客观标准。

（一）事实性标准

事实性标准是教学标准（或职业标准）中规定和描述的能力要求，借此可以对学习结果进行客观评价。事实性标准的优点是：评价结果与教学标准要求的关系清晰可见，可以明确判断是否达到了最低要求，从而权威性地证明达到了某一特定要求（如技能等级）。事实性标准可帮助学生进行目标导向和自我管理式学习，但是无法让教师了解学习难度，对学生的进步也不敏感。另外，在实践中经常很难严格和客观定义什么是优秀、良好或者及格（如很多服务和技术过程无法精确评分）。如果学校对学生的成绩分布有要求（正态分布）时，教师还必须根据下面要谈到的社会性标准对考试分数进行修订，这不但会造成结构性困难，而且很容易引发矛盾，为教学设计带来阻碍。

（二）社会性标准

社会性标准是与他人对比时学生表现出来的绩效差异，即学生在一个团体中的相对水平。社会性标准的最大优点是能够识别表现相对较好和较弱的学生，确定能力差异，为开展个性化教学提供基础。其局限性是：可比性只是在特定的团体内有效（班级内部标准）。由于看不到全体成员的共同变化，所以所确定的成绩有可能掩盖全体成员的共同努力或相同的退步，掩盖事实性标准的实现程度，以至失去标准本来的理性和可信度。而且社会性标准无法精确确定考生的分数，这有可能对其学习动机造成消极的影响。

^① Rheinberg, F. Bezugsnormen und schulische Leistungsbeurteilung[C]//Weinert, F.F. Leistungsmessungen in Schulen. Weinheim: Beltz, 2001: 59-71.

（三）个性化标准

个性化标准是与自己的过去进行比较，由此追踪学生的学习和成长过程。个性化标准的优点是：无论是好的还是差的学生都能确定自己的学习状态，自身努力与学习成就之间的关系清晰可见，不同阶段的绩效波动能够可视化，这有利于提高学习者的学习动机。采用个性化标准，后进生不会受到优秀同学的打击，而优秀学生也可以与“过去的自己”相比较。但是个性化标准弱化了学生之间确实存在的差异，使学生失去客观评价自己的信息源，不适合评价外部有特定要求的领域（如入学考试和职业资格考试）。

综上所述，每种评价标准都有其优势和“盲点”，不可能满足学业评价的所有要求，教师在实践中需要利用多种标准并相互补充，学生也应学会利用各种标准进行自我评价。学业评价标准的理想状态是，个性化标准发挥主导作用，但又不忽略其他标准提供的信息；事实性标准与个性化标准相结合，是未来职业教育学业评价发展的趋势。

四、学业评价方法分类

根据不同的背景、需求和目标，学业评价会采用不同的方案，这涉及不同的设计理念、数据收集、分析和结果解释的方法。学业评价有多种分类方法^①，本文不进行分类学研究，只进行简单的类别阐述，为后面的讨论奠定基础。

（一）量化评价

19世纪中叶人们开始进行大规模的学业成就测验，其标志性事件是赖斯（J. Rice）对33 000名学生进行的拼写能力调查。20世纪30年代，“评价之父”泰勒（R.W. Tyler）采用实验设计模式对教学改革成果进行比较分析，量化评价成为学校教育最重要的学业评价方法。职业教育中采用的量化评价有能力倾向测试、学习成就测试、人格测试和能力测评等多种方法。其中，能力倾向测试对个人的潜在能力进行评价，预测个体能力

^① 李坤崇. 学业评价：多种评价工具的设计及应用[M]. 上海：华东师范大学出版社，2016.



未来发展倾向；学习成就测试对个体的学业成就进行测试，典型的如考夫曼教育成就测验-II (KTEA-II)^①；人格测试通过特定测试量表（如明尼苏达多相人格调查表、艾森克人格问卷等）对个体的心理特征进行测试，如兴趣、动机、态度、气质和性格等^②；能力测评方法很多，既有单项能力测评方法，也有综合能力测评方法，如BCI方法通过性格问卷对个人能力、自我组织和自我控制能力进行诊断评价^③，基于等级描述分类的混合式能力评价方法SOLO (structure of the observed learning outcome) 通过解答问题的表现特征对人的思维发展进行可视化分析^④。

量化评价可以避免考官个人主观因素对评价结果的影响，对促成目标达成有较好效果，因此在学业评价中发挥着重要作用。但是科学的量化评价模型很难建立，这极大地影响了评价结果的说服力，而且这些方法在过程性评价中的使用效果往往也不理想，采用互动性更强的质性评价更容易深入理解和解释被观察的现象（详见第二章）。

（二）质性评价

人力资源管理最先开始采用基于胜任特征模型的质性评价方法，如“工作胜任力测评法” (job assessment method, JACM) 和“行为锚定修正法” (modified bars method) 等。这些方法建有能力模型，但是没有量化工具和流程，致使其操作过于灵活。德国开发实施了数十种职业能力测评方法，其中有很多是质性方法，如行动导向测评方法、单项技能评价方法和能力总结与汇总方法 (competence balance) 等。我国一些职业院校采用类似方法进行形成性评价，如北京市商业学校以职业素养学分和证明、证章评价方式，记录学生在思想品德、技术技能、文化艺术、体育科技等方面取得的成绩，并在此基础上给学生颁发“职业素养护照”^⑤。苹果公司 A+ 雏鹰计划在其试点院校校企合作学徒培

① 格雷戈里. 心理测量历史、原理及应用 [M]. 北京: 机械工业出版社, 2013: 194-195.

② 戴海崎, 张峰, 陈雪枫. 心理与教育测量 [M]. 广州: 暨南大学出版社, 2011: 13-14.

③ Bambeck, J.J. Bambeck-Competence-Instrument[C]//Erpenbeck, J. & von Rosenstiel L. Handbuch Kompetenzmessung. Stuttgart: Schäffer-Poeschel, 2007: 3-22.

④ 李佳, 吴维宁. SOLO 分类理论及其教学评价观 [J]. 教育评价与测量, 2009(2): 16-19.

⑤ 高靓. 职业素养护照为学生职业发展对标对表. 中国教育新闻网 [EB/OL]. http://www.jyb.cn/rmtzcg/xwy/wzxw/201912/t20191230_285442.html, 2019-12-30.

养中建立“学习档案”(portfolios)^①,学习者有意识地搜集证据,记录和反省自己在某一领域的进步,用可靠的证据记录取得的进步并获得绩点。

质性评价的主要缺陷是结果的可比性较差。由于对学生的行为描述、分类、选定能力指标和确定界别的难度较大,所以此类方法对评价人员的专业要求也很高。此外,如果被评价者过于谦虚、夸张或不诚实,评价结果的可信度也会降低。

(三) 混合式评价

职业能力是个体从事一个职业时的心理认知特性,波兰尼“隐性知识”理论被广泛接受,间接否定了按照行为主义理念采用纯量化方法进行能力测量的合理性^②,量化与质性相结合的混合式评价成为发展方向,国际有影响的职业能力评价方法均据此建立技能模型和测评模型。

20世纪末开始的PISA等普通教育大规模教育评价项目也引发了全球对职业教育质量评价方式的讨论。讨论得到的一个共识是:职业教育质量控制应实现从“投入导向”向“结果导向”的范式转变,而作为学业成就标志的职业能力是职业教育最重要的“结果”。COMET(competence measurement)职业能力测评是世界上第一个真正实施的大规模职业能力诊断工具,它基于“从初学者到专家”的能力发展逻辑^③建立包含“级别”“内容”“行动”三个维度的跨职业能力模型,采用开放性测试题目,确认被测学生的职业能力发展级别,即名义能力、功能性能力、过程性能力和设计能力^④。教育部督导组制定的《高等职业教育专业评价试点方案(草案)》采用了COMET方案,对三个代表性省份现代物流管理等四个专业的高职学生进行试验性的能力测评。结果显示,COMET能力测评可以较好地确认不同院校学生的职业能力发展水平,其开放性测试题目不仅被企业认可,也广受学生欢迎。

COMET能力测评不仅适合宏观层面和中观层面的学生学业评价,对不同院校的教学质量进行比较,也适合微观层面的教学评价,对教师的教学设计提供指导。COMET

① McMullan M. et al. Portfolios and Assessment of Competence[J]. J Adv Nurs, 2003, 41(3): 283-294.

② 石中英. 波兰尼的知识理论及其教育意义[J]. 华东师范大学学报(教科版), 2001(2): 36-45.

③ Dreyfus H.L. & Dreyfus S.E. What artificial experts can and cannot do [J]. AI & society, 1992, 6(1): 18-26.

④ 赵志群, 劳耐尔, 等. COMET职业能力测评方法手册[M]. 北京: 高等教育出版社, 2018.



职业技能评价方法研究

还被工业和信息化部及世界技能大赛与“一带一路”技能大赛等不同的技能评价工作采纳。在设计能力测评实施方案时需要特别考虑的是，哪些评价手段既能获得大量深入和有效的数据，又不会引发过多的经费和时间投入。在此需要解决的关键问题，一是关于产出和投入的因果关系分析，二是具体评价工具的开发，包括测试题目类型的研究等。

（四）诊断性评价

诊断性评价在我国的相关实践中具有重要的意义。“诊断”的概念起源于医学，本义是通过特定的工具检查病人的病症，并有针对性地开具处方。布鲁姆（B. Bloom）最先在教育领域提出“诊断性评价”概念，即在学习开始之前对学习者现有水平进行评价，强调预防性的“事前行为”^①。诊断性评价是内部评价方法，其最大特点是采用“协商模式”，即相关人员协商裁定和解决相关问题，从而提高评价结果使用的可能性。诊断不是简单的现状描述，而是通过获得足够深入的信息，揭示学习中存在的问题及其成因，并通过制度化的措施形成持续改进的机制。

进入 21 世纪，诊断成为引领学校建构反思意识、发现问题、提高效能、促进自主发展的新型教育技术，特别是教育部通过促进建立职业院校“教学诊改”机制的行政措施，使这一思想得到了普及。值得注意的是，诊断不仅仅是收集数据，更重要的是使利益相关者都参与到诊断过程中，共同发现问题、寻找解决策略并实施改进。学业成就的诊断性评价是采用量化或质性工具，根据职业教育利益相关者的共同诉求（“会诊”）对学生在学习取得的成果和存在的问题（“病情”）进行多维度分析，确认和总结原因（确诊“病因”），并提出解决策略（“开处方”），从而提高学习者的学习能力，以取得更大成功（“健康”）。

五、“1+X”证书对职业技能评价技术的要求

国务院颁布实施的《国家职业教育改革实施方案》提出在职业院校和应用型本科高校启动“学历证书+职业技能等级证书”（即“1+X”证书）试点工作，这是我国在双

^① 布卢姆. 教育评价 [M]. 上海: 华东师范大学出版社, 1987: 10.