

高职课程学业表现性评价的任务框架设计与评分量规设计

殷明, 季琼

[摘要] 高等职业教育侧重于培养职业性、应用型的高素质技术技能人才,在课程教学中注重能力本位与工作任务情境,传授“功能性知识”,课程学业更适合采用表现性评价以衡量学生的认知思维与推理技能,以及运用知识解决现实、有意义的问题的能力。表现性评价设计的核心包括评价任务设计与评分量规设计。评价任务设计框架包括主题内容、行动或实践、情境、意义,以有效地引发学生“证明其习得功能性知识”的表现。评分量规设计框架包括评分指标、指标权重、质量水平层次、水平分值、标准描述,可作为表现性评价设计具体步骤的参考,最终促进高职课程教学实践更好实现育人效果。

[关键词] 高职课程;学业表现性评价;任务框架设计;评分量规设计

[基金项目] 全国关务职业教育教学指导委员会2023年度教科研项目重点课题“基于能力本位的高职关务与外贸服务专业课程学业表现性评价研究”(项目编号:2023003,主持人:季琼)

[作者简介] 殷明,博士,广东白云学院,副教授;季琼,硕士,北京劳动保障职业学院,教授。

中图分类号:G710 文献标识码:A 文章编号:1004-9290(2024)0011-0076-10

引言

学业评价是指以教学目标为依据,运用恰当的、有效的工具和途径,系统地收集学生在各门学科教学和自学的影响下认知行为上的变化信息和证据,并对学生的知识和能力水平进行价值判断的过程^[1]。课程是高等职业教育人才培养的核心环节,课程学业评价是检验课程目标达成度的手段,也是毕业要求达成度评价的基础。2018年,教育部提出高等职业教育要打造“金课”,强调培养学生解决复杂问题的综合能力与高级思维。2020年10月,中共中央、国务院《深化新时代教育评价改革总体方案》提出“改革学生评价,严格学业标准,完善过程性考核与结果性考核有机结合的学业考评制度”,对课程学业评价的有效实施提出了更高的要求。然而,传统测验评价方式侧重于对学生获得多少结果性知识的评价,显然无法检查学生的高阶

能力(可迁移能力、批判思维和问题解决等)^[2],因此在高职课程中引入表现性评价(Performance Assessment)作为一种替代性评价,并重点分析评价任务框架与评分量规设计思路和案例,以达到促进高职院校学业表现性评价中知识与技能联动的效果^[3]。

一、表现性评价的内涵及特征

作为标准化测试的替代评价方法,表现性评价并不是某一种具体的评价方法,而是一系列评价方法的统称,它们具有共同主线,主要包括:(1)学生需要使用高级思维或解决问题的技能;(2)任务是在真实环境或仿真模拟中设置的有意义的挑战;(3)评价对象包括产品、过程与认知行为^[4]。因此表现性评价也并不是仅仅评价学生外显的动作行为表现,也包含了在语言、逻辑、艺术、社交方面的表现,还包括解决问题的过程与结果。这里的“表现”也不是日常表现

或完成某一简单任务的表现,而是在新情境中综合运用所学知识、技能的表现,同时还涉及到高阶思维。而“完成任务”本身并不是评价的对象,而是“表现”的载体,真正被评价的是所学到的知识、技能应用的熟练程度。

关于表现性评价,已有众多研究学者开展了相关问题的研究。例如,周文叶从评价的过程视角将表现性评价定义为“在尽量合乎真实的情境中,运用评分规则对学生完成复杂任务的过程表现或结果做出判断”,并解析表现性评价的内涵为真实情境中知识和技能的运用、学生必须自己创造出问题解决方法(答案)或通过自己的行为表现来证明自己的学习过程和结果、评价者根据事先设计好的评分规则做出判断^[5]。对于如何激发学生将所学知识、技能加以应用,这就需要提供一基于现实世界的复杂的任务情境,学生在情境中根据任务要求建构他们的反应。这一建构反应过程中,学生必须应用所学的多项知识、技能,且反应的结果是开放性的,并不具有唯一的标准答案。但同时建构反应的过程与结果进行评价往往带有较大的主观性,因此必须预设表现的质量标准,且该质量标准应该映射所预期展示的多项知识、技能,即标准含有多项评价指标。从评价是为了促进学习这一理念出发,预设的评价标准一般也会事先向学生公开,作为学生建构反应的线索或自我评价的参考。

二、高职课程学业评价采用表现性评价的必要性

课程设计的一个核心是确保所选择的评价方法最适合课程的内容、教学方法与学习结果,特别是在高职院校课程中^[6]。

(一)人才培养定位决定高职课程传授“功能性知识”

高等职业教育培养高素质技能型人才、技术技能型人才或者应用型人才,这一人才培养定位包含了“技能与应用”要素。技能与应用的主体是学生本人,客体是学生所学的知识主题内容,如理论、原理、规则、方法、工具、技术等。

如果学生在课程结束时只是能够将这些知识陈述出来,那就不是技能与应用。真正的技能或应用是学生能够在理解的基础上、在特定情境中执行或应用这些知识,即赋予这些知识以实际的情境性功能或现实意义。此时,这些知识就成为“功能性知识”——根据对事物的理解,以行动或实践来体现知识^[7]。

(二)学习领域课程开发模式决定高职课程传授“功能性知识”

高等职业教育人才培养目标决定了高职院校学习领域课程开发的模式。该课程模式下,学习领域是以一个职业的典型工作任务(是职业行动中的具体工作领域,也称为“职业行动领域”)为基础的专业教学单元,相当于一门课程,由一个或若干学习任务组成。学习任务的实质是“学习与工作任务”,即“内容是工作的学习任务”或“用于学习的工作任务”^[8]。因此,学习领域课程传授的知识侧重于“如何工作”——这是需要以行动或实践来体现的,属于“功能性知识”。

(三)学生通过表现性评价证明其习得“功能性知识”

“功能性知识”是学生以行动或实践来展现其对知识的理解。这一定义直接表明对功能性知识的评价是一种“表现性评价”。

首先,表现性评价要求学生自行构建答案、生产(创作)产品(方案或作品)、执行活动,相比较传统评价的“从记忆中提取一段信息默写出来,或从提供的选项中识别出一个潜在的正确答案”等方法,更加能够激发学生以行动或实践来展现其对知识的理解。

其次,表现性评价必须蕴含学生对新知的理解,而不是仅仅依赖学生原有知识与经验而完成任务。否则,评价就偏向于行动或结果本身,而不是功能性知识。例如,对高职关务与外贸服务专业学生的进出口商品归类技巧学习的评价就不是查询进出口商品HS编码这个行动或者查询的结果(即使学生通过百度查询得到正确的结果),而是在查询商品HS编码的过程中是否应用了所学的商品归类技巧。行动是让

学生展现其对新知的掌握手段或载体。因此评价标准应围绕新知的理解或应用,而非仅仅围绕行动或结果本身。

再次,表现性评价一般要求学生在特定的情境中完成任务,以展现知识的功能或作用。在特定的情境中,被理解(或应用)的知识与行动或实践结合在一起,创造出一个新的功能,在情境中体现出它的作用或意义。没有情境,功能的意义就无从谈起。没有情境,学生所完成的行动只不过是交给老师的一份作业。这样的知识虽然学生表现出来了,但学生可能会迷惑这些知识学到之后的用处。

综上所述,高职课程传授的内容偏向“功能性知识”,要求学生以行动或实践来展现其对知识的理解,这与表现性评价的特征是一致的。与传统的纸笔测验或者简单的去情境化操作测试相比,高职课程采用表现性评价,更加贴合高职课程学业特点,同时也更加能够与高职课程学业目标水平保持匹配一致。

三、高职课程学业表现性评价任务设计框架

高等职业教育参与者实施表现性评价面临的主要问题是表现性评价的设计,高等职业教育似乎缺乏一个共享的标准或一套规则来指导表现性评价^[9]。周文叶认为完整的表现性评价由目标、任务、评分规则三部分组成^[10]。目标是设计表现性评价的指引和指导原则,表现性评价需要与课程目标对齐,但课程目标本身并不是表现性评价设计的主体内容。因此表现性评价设计的核心部分为评价任务的设计与评分规则的设计(在此两项设计中必须满足与课程目标对齐)。

评价任务应该能够清晰地映射课程预期学习成果^[11]。一个恰当的评价任务应该能够告诉教师某一个学生的预期学习成果完成得怎样,其作用是引发学生“证明其习得功能性知识”的表现,并明确所涵盖的内容标准。一个完整的表现性评价任务包括四个要素:主题内容、行动(或实践)、情境、意义。其中,行动(或实践)又包括四个子要素,分别是行动的要求、条件、结

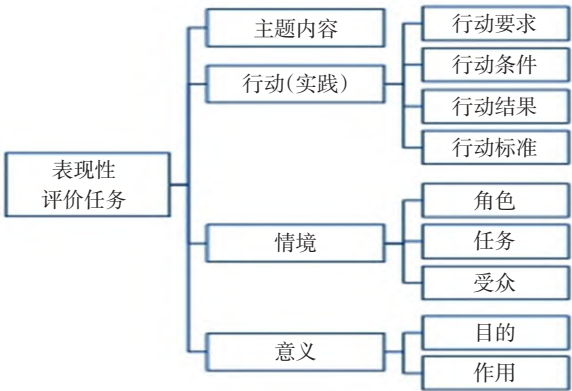


图1 高职课程表现性评价任务设计框架^[12]

果、标准;情境也包括三个子要素,分别是角色、任务、受众;意义则包括两个子要素:目的与作用(图1)。

(一)主题内容

“功能性知识”需学生以行动或实践来展现其对知识的理解。主题内容即其中的“知识”。它既是表现性评价的动因(但它不是评价的对象,评价的对象是“对知识的理解”),也是评价标准的来源。因此在一个表现性评价任务中,主题内容应至少在3个地方呈现,即评价任务的标题名称、评价任务的学生行动要求以及评价的标准。

(二)行动或实践

表现性评价任务应明确学生为证明其习得主题内容所表现出的行动或实践。具体包括在行动的要求、条件、结果、标准这4个子要素上的清晰界定。

1. 行动的要求

学生具体要做出哪些行动(以证明其对主题内容的理解或新知识的建构)应参照预期学习成果中的表述(尤其是其中所使用的行动动词),并匹配至相关的教育目标领域与水平层次。为促进学生围绕“对主题内容的理解或新知识的建构”做出行动,在行动的要求中往往结合主题内容给定行动的框架作为指引。

2. 行动的条件

这里所指的行动是学业评价的行动(大多数是终结性评价),而非平时练习的行动,因此需要从评价特征的视角对学生的行动限定条

件。比如,时间限制、重做限制、完成的独立性限制、使用工具的限制等。

3. 行动的结果

学生行动的结果对应于被评价的表现,此时结果既可以是行动的产物(如学生设计的一份方案),也可以是行动本身(即过程,如学生展示的一段舞蹈)。在这个意义上,行动的结果更接近于被评价的表现的载体。作为载体,它具有可保存性,这就满足了教育的质量保证体系要求。在某些情况下,教师还需要学生提交行动的过程以确认学生对主题内容的真实理解。此时的载体一般采用历程档(Portfolio)的形式,它内含阶段性的成果、草案以及过程性的描述、照片、视频等。

4. 表现的标准

表现的标准也就是评价标准。学生根据这一标准来控制其行动的程度或质量,并进行表现的自我评价。教师也是根据这一标准对学生的表现进行评价。在高职课程学业表现性评价任务中,评价标准主要源于相关职业标准,但需要注意课程学业评价任务情境与职业真实情境的差异。比如,在高职“通关实务”课程中,有一项学生模拟代理报关服务人员与委托人(进出口货物收发货人)洽谈代理报关业务的评价任务,任务情境的设计是教师与企业协同完成的结构化情境,与职场中多变复杂的真实情境相比,对学生表现的要求相对会低一些。

(三)情境

表现性评价任务应明确学生行动或实践的具体情境。这就意味着,学生的行动或实践展现的是主题内容在具体情境中的理解或应用。情境包括3个子要素:任务、角色、受众。

1. 情境任务

学生在情境中的行动,不是孤立的或机械的行动,而是基于一个具体的任务。学生通过完成任务的结果或完成任务的行动过程来做出有待评价的“表现”。任务的典型性可以提高表现性评价的效度。在高等职业教育领域,任务的来源可以是有代表性的工作任务或者是职业

资格证书考核中的任务,也可以是任课教师围绕主题内容自行构建的任务。

2. 情境角色

情境角色是学生作为行动者在情境任务中的角色。这一角色应该与职场或社会生活中的真实角色相接近。只不过因任务设置的不同,学生在表现性评价任务中的角色活动与真实角色活动可能会有些差异。但最起码要让学生明白该角色的意义并在任务行动中体验该角色。

3. 情境受众

受众是指学生在情境任务中是服务于或作用于哪个角色,或者完成此任务后传递给哪个工作角色。一方面,受众强调让学生理解其任务并非孤立;另一方面,有助于培养学生换位思考素养。同时,情境受众与情境角色一起,作为培养或考察学生在角色担当过程中自主性与责任心的重要手段。

(四)意义

一个有意义的学习成果需要匹配有意义的表现性评价任务。所谓有意义的表现性评价任务,是指表现性评价任务对学生而言并非仅仅是一个作业或考试,而是关注其内在的意义。这体现在两个方面,一是目的——即该评价任务的目的是考察学生对特定主题内容的理解,或者说是考察学生习得什么样的能力,也就是通过评价任务要求学生所做出特定的行动或者展示以证明其习得了特定的能力;二是作用——该作用并非在学生能力展现方面的作用,而是评价任务在现实世界的作用,这涉及到另一个概念“真实性评价”。比如,在高职“通关实务”课程中,学生模拟代理报关服务人员与委托人洽谈代理报关业务,其中涉及到“对客户违法要求的拒绝”。这绝不仅仅是为了刺激学生表现出守法的素养,而是考虑到这是在职场较常经历的现实情境。

四、高职课程学业表现性评价的评分量规设计

表现性评价的一个重要特征是“对学生的行为表现采用多项评量指标,并在评量前对每

项指标的具体标准进行严格清晰的界定,由测量人员根据各指标标准,主观性地对学生的行为表现进行综合评估”^[13]。由此形成的评价工具被称为“评分量规”(Scoring Rubrics,也称为“评分量表”或“评分规则”)。评分量规由一个固定的量表和一系列描述量表上每个分数的表现的特征组成^[14]。它在形式上有点类似于教育领域竞赛或选拔所用的评分表,不同的是每个指标的标准描述并不只是最优表现,而是分等级(层级)的多个表现描述。评分量规可以分解为5个要素:评分指标、指标权重、评分标尺、标尺刻度、标准描述。

(一)评分指标

评分指标也就是评分量规的“量”的维度。即对成果表现中的哪些要素进行评分,此阶段要素既可以是完成评价任务而需要应用的知识点或技能点,也可以是掌握的质量点,或者二者合一,必要时设计二级评分指标,对要素细项单独评分。选择哪些知识点、技能点或质量要素作为评分指标,取决于表现性任务所对应的能力点,亦即表现性任务所对应的课程目标所支撑的专业学习成果(PLOs),这涉及到对课程预期学习成果或评价任务的具体分析^[15]。

实际上,这些要素指标并不是在设计评分量规时才予以考虑的,而是在设计表现性评价任务时就需要考虑,因为这些要素需要与学生完成任务所提交的产品(也包括部分完成任务的过程)相匹配,也即在学生完成表现性评价任务所提交的产品(或过程)中,应能“可视化”地采集到与要素指标相关联的信息,否则即使手握一份完美的评分量规,也难以对成果表现进行有效评分。在教学实操中,可以应用“表现性评价定位框架(图2)”作为逻辑起点来理清思路。该框架包含以下两个维度。

1. 评价任务的定位,即评价任务是用来评价什么

在能力本位教育中,表现性评价任务基本上都是用来评价能力的。具体而言,一般有两种不同的定位导向。一种导向是评价特定的能

以高职“通关实务”课程为例

表现性评价任务:基于给定情境,模拟完成代理报关业务的洽谈并签订代理报关服务合同

所对应的课程目标:能合法诚信地商定代理报关业务

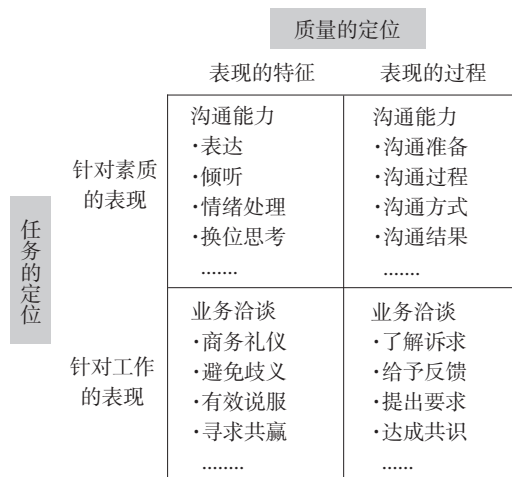


图2 高职课程表现性评价定位框架

力(如某一核心能力或关键能力,为区别于广义的能力,在图1中使用了“素质”这一术语),当然是基于表现性任务的执行来评价能力,也即将表现性任务作为评价能力的情境载体;另一种导向是评价特定的任务本身,这在基于工作过程系统化所开发的课程中会较多见到,即评价岗位代表性工作任务的执行与完成情况。此时,能力是蕴含在工作任务之中的。因此在设计表现性任务的评分量规的评分指标时,首先确定该表现性任务的定位是属于哪一种导向。导向的不同,决定了评分指标的不同。

2. 表现质量的定位,即按照什么思路对表现的质量进行分解

一般也有两种分解思路。一种是基于表现的特征进行质量分解,即通过哪几个表现特征来评价表现质量水平的高低;另一种是基于表现的过程进行质量分解,即根据表现中各个步骤过程或各组成部分的表现来评价表现质量水平的高低。

上述两个维度的不同导向,交叉组合成一个表现性评价定位的2×2矩阵框架。图2以高职“通关实务”课程的一个课程目标“能合法诚信地商定代理报关业务”中所涉及的“沟通能力”为例,对此框架进行分析。

首先,在“任务的定位”维度,评价任务可以直接面向沟通能力,而较少考虑任务本身的完成情况,因此在图2的上半框中并没有界定具体采用什么任务来评价“沟通能力”;评价任务也可以是具体明确地界定为“业务洽谈”任务(图2的下半框),此时评价的重心就不再是沟通能力,而是业务洽谈能力,沟通能力是内含于业务洽谈能力之中的。

其次,在“质量的定位”维度,可以根据表现的特征来分解质量,其中对于面向素质(沟通能力)的表现,其表现质量特征可以分解为表达、倾听、情绪处理、换位思考等特征(图2左上框);对于面向工作任务(业务洽谈)的表现,其表现质量特征可以分解为商务礼仪、避免歧义、有效说服、寻求共赢等特征(图2左下框)。在此维度,也可以根据表现的过程来分解质量,其中对于面向素质(沟通能力)的表现,其表现过程可以分解为沟通准备、沟通过程、沟通方式、沟通结果等环节(图2右上框);对于面向工作任务(业务洽谈)的表现,其表现过程可以分解为了解诉求、给予反馈、提出要求、达成共识等环节(图2右下框)。在评价时,则根据上述各个环节的表现质量进行评价。从本示例中得到以下几点启示。

一是评价任务定位的不同(面向素质抑或面向工作任务),会导致表现质量特征的主体不同。面向素质的评价,表现质量特征的主体是素质;而面向工作任务的评价,表现质量特征的主体是工作任务。主体的不同,导致了表现质量特征的不同。其中,最明显的差异是面向素质的评价侧重于素质的技能技巧;而面向工作任务的评价侧重于工作任务本身完成的质量,此时相关素质是其特质体系中的一部分(即其中一个或两个评分指标),尽管这一两个评分指标的权重系数可能会比较高。

二是评价任务定位的不同(面向素质抑或面向工作任务),在表现过程分解方面的差异并不大(除了表述方面,一个围绕素质,另一个围绕任务情境来展开表述)。其原因在于从过程角度分析,素质(作为基于任务的能力)和工作任务

(以能力为核心的任务)基本上是同一主体。

三是基于表现过程的质量定位,从范畴层级而言,比基于表现特征的质量定位更高一层。也就是可以在表现过程的各个环节(即一级指标)下,再设置相关的表现特征作为二级指标。尽管这一做法相对复杂,但其优势在于更加符合工作过程系统化的课程设计理念以及更加符合循序渐进的教学规律。

(二)指标权重

指标权重是指评分量规中的每个要素指标(包括二级指标)在整个指标体系中所占的权重,一般用百分比权重系数来表示,实际上它代表了各要素指标(包括二级指标)在整个表现性任务评分中的成绩比重(合计为100%)。权重系数的确定,一般可以参考3个因子:一是该表现性评价任务所对应的课程目标所支撑PLOs的权重系数(或贡献度系数);二是各评分指标相对于完成表现性任务而言的重要性,或者说评判表现性任务成果优劣的重要性;三是各评分指标的难度或挑战性(如思维的高阶性、心因动作技能的复杂度或熟练度、情感的内化程度等),它在一定程度上代表了学生完成该评分指标所对应的任务所需要的时间与精力。图3以高职“通关实务”课程为例,展示了评分指标权重设计的思维过程。

(三)表现质量水平层次(评分标尺)

评分量规的一个与众不同的特征就是并不只是界定每一评分指标上表现的最优质量,而是界定多个不同层次等级的质量水平。从理论上讲,层次等级分得越多,评分可以做到越细致,但不同等次等级之间区分度越难以控制,评分量规的设计要求也越高。一般而言,评分量规中每一评分指标界定3~5个水平质量层次,相当于构建了一系列3~5个刻度的评分标尺,每一刻度对应着不同的表现质量水平。丹奈尔·史蒂文斯建议刚接触评分量表时,一般使用3等级刻度,熟练后往往可以扩展为5等级刻度^[16]。但笔者在实践中发现,5等级刻度的评分标尺在描述每个等级的质量标准时难度较大,每个等级

以高职“通关实务”课程为例
表现性评价任务:基于给定情境,模拟完成代理报关业务的洽谈并签订代理报关服务合同
所对应的课程目标:能合法诚信地商定代理报关业务

支撑的 PLOs	支撑系数	评价要素	重要性	难度	权重
PL01.2 核心价值观	0.1	合法诚信	5	1	15%
PL02.1 关务知识应用	0.3	条款齐全	4	2	20%
		内容准确	5	4	45%
PL03.2 商业沟通能力	0.1	表达与理解	4	2	15%
		谈判技巧	1	4	5%

备注:1.示例中“支撑系数”已经过标准化处理;
2.“重要性”与“难度”均按 1~5 分赋值。其中,1 表示重要性最小,5 表示重要性最大。

图3 评分权重设计示例

之间的区分度较难控制,因此,建议采用 4 等级刻度即可。

评分标尺的等级赋名。赋名是指为评分标尺上每个质量等级赋名,一方面,可以明确质量等级的高低;另一方面,也用文字概括了每个质量等级的达标状态,因此一般是以描述状态的形容词进行赋名。Huba(胡巴)、Freed(弗里德)罗列了一些常用的描述表现质量水平层级的形容词(图 4)^[17]。

Huba & Freed(2000)常用的描述表现水平层级及形容词:
· 优秀、合格、部分合格、不合格(NSF 综合工程教育联盟,1997)
· 典范、娴熟、勉强合格、不合格
· 高级、中高级、中级、初级(美国外语教学委员会,1986)
· 优异、娴熟、中级、初级(Goteher,1997)
· 精通、普通、有待进步、入门(教育学院,1997)

图4 常用的描述表现质量水平层级及形容词赋名

与评分标尺等级赋名相关联的另一个决策是确定“及格”临界点的刻度,即哪一个表现质量的水平层级为“及格”水平。从图 4 可以看出,在四等级评分标尺中,大部分为有三个等级为及格及以上水平,只有一个为不及格水平,这与我们常说的优良中差相一致。但笔者在实践中发现这并不好借鉴。其原因在于我们一般设定为 60 分为及格线(百分制),这样一来,0~100 分的跨度中,0~59 分的区间只有一个水平等级,而 60~100 分的区间竟有三个水平等级,造成了评

分标尺刻度的较大不均等,以及 60~100 分区间内设置三个水平等级难以体现区分度。因此笔者建议在以 60 分为及格线(百分制)的情况下,四等级评分标尺中设置两个水平等级为及格线上,两个水平等级在及格线下。此时,对评分标尺刻度赋名就需要相应调整以保持一致,如优良、及格、较差、差(图 5)。

(四)评分标尺的刻度赋值(标尺刻度)

为了后续便于计算,需要为评分标尺上每个质量等级赋一个数值,如 1、2、3、4,代表从低到高的四个质量等级(从评分成绩计算角度出发,等级越高则数值越大),这也是最常使用的

以高职“通关实务”课程为例
表现性评价任务:基于给定情境,模拟完成代理报关业务的洽谈并签订代理报关服务合同
所对应的课程目标:能合法诚信地商定代理报关业务

评分指标	优良 4	及格 3	较差 2	差 1	指标 权重	评分 乘数	教师 评分	百分制 分数
合法诚信	※				15	15/4	4	15
条款齐全	※				20	20/4	4	20
内容准确		※			45	45/4	3	33.75
表达与理解		※			15	15/4	3	11.25
谈判技巧			※		5	5/4	2	2.5
合计					100	100/4	—	82.5

备注:图例中指标“表达与理解”的百分制分数 11.25 是用评分乘数 15/4=3.75 乘以教师评分值 3 而得来,其中评分乘数中的 15 源于该指标的权重系数,4 源于该指标的最高表现质量等级赋值为 4。

图5 评分量规的评分等级刻度及评分计算

赋值方法(如果评分标尺设置了五个等级,则赋值为1、2、3、4、5)。需要说明的是,这里的赋值(1、2、3、4)并不代表学生在相应评分指标上的实际成绩,只是作为实际成绩计算过程中的一个“中间代值”(图5)。

评分标尺的质量水平等级数与各评分指标权重值一起构成了各评分指标的“评分乘数”,再结合评分标尺的刻度值(学生的实际表现质量符合评分标尺的哪个水平等级,就取该水平等级刻度值)就可以和方便地计算出成绩(图5),这也是评分量规作为快捷评价工具的体现之一。但需要提醒,与目前百分制评分的100个等级相比,评分量规采用四个或五个等级评分的颗粒度实际上差了很多。但这同时也表明评分量规这一评价工具在适用场合上更偏向于难以精确界定到1分或2分的评价任务。也就是说,如果某一表现性评价任务的评分标准能够很明确地界定怎样的情况下(比如,每错1处或者每多或少一个量值)扣1分或扣2分,那就不需要采用评分量规,而是直接界定扣分细则。

(五)各质量等级的标准描述

各评分指标的不同表现水平的质量标准如何进行描述以对应不同表现水平的明确区分度?在回应这个问题之前,必须先需要考虑根据什么原则或方法来对任务表现水平进行分层分级,否则各层级标准描述就具有较大的随意性。笔者在实践中归纳了三类原则或方法,具体如下。

1. 原则或方法一:教育目标分类水平层级
布鲁姆教育目标分类体系^[18]或彼格斯SOLO(可观察的学习成果结构)体系中^[19],已经界定了在某一领域教育目标的不同水平层次。按照这些模型来对任务表现水平进行分层分级,就是

教育目标分类中的低阶目标对应任务表现的低层次水平,教育目标分类中的高阶目标对应任务表现的高层次水平。比如,在高职“通关实务”课程中的“基于给定情境,模拟完成代理报关业务的洽谈并签订代理报关服务合同”这一表现性评价任务中,其中一个子任务是拟定一份报关服务合同,这涉及到认知领域目标。学生较低的表现是直接套用一份报关服务合同范本,这对应于简单的“理解”层次;相对好一点的表現是从多份报告服务合同范本中选择适当内容加以组织,这对应于“分析”或“综合”层次;再好一点的表现是能够根据给定的情境,对报关服务合同范本中部分内容进行修订或增补新的内容条款,这对应于“评价”或“创造”层次。由此,即是按照教育目标分类水平层级来确定任务表现的不同等级。

另外,在实践中还有一个捷径方法,即先根据课程学习成果的要求来确定最佳表现所对应的教育目标水平层级,然后通过分析该教育目标水平层级的要求“是什么”与“不是什么”来进行任务表现等级区分(图6)。

2. 原则或方法二:最佳实践与常见错误
- 丹奈尔·史蒂文斯在《评价量表:快捷有效的教学评价工具》提出建构评分量规四步骤中的第二步即为列举学生任务表现的最高期望水平,将之整理后(如分组、合并)作为评分量规相关评分指标的最高质量等级标准描述^[20]。与此相对应,也可以列举学生在完成任务中的常见错误,将之整理后作为评分量规相关评分指标的较低质量等级标准描述(图7、图8)。
3. 原则或方法三:关键表现
- 一般通过最佳实践与常见错误的列举分

目标水平层次	是什么	不是什么
知识与认知 创造	是经过学生自己的构思与综合,基于对知识的理解	不是直接照搬
知识与认知 应用	已学概念在具体情境中的套用,包含对概念的理解	不只是提取概念的名称甚至没有应用概念
态度与素养 接受	是初步的理解与认同	不是形成习惯也不是社会性赞许

图6 教育目标水平层级的要求分析示例

以高职“通关实务”课程为例

表现性评价任务:基于给定情境,模拟完成代理报关业务的洽谈并签订代理报关服务合同

所对应的课程目标:能合法诚信地商定代理报关业务

评分指标	最佳实践(示例)	常见错误(示例)
合法诚信	· 对客户提出的违法要求予以拒绝	· 为达成业务而向客户作出不能实现的承诺
条款齐全	· 有详细明确的服务内容 · 收费条款和结算方式 · 委托双方的权利和义务 · 违约责任与争议解决 · 单证附件	· 从客户处获取的通关单证有重大缺失
内容准确	· 无明显错误	· HS 编码错误 · 监管方式误用
表达与理解	· 对己方的服务承诺作出清晰的说明	· 在沟通中对客户诉求产生理解偏差
谈判技巧	· 在商谈中应用“说服与共赢”技巧	

图7 任务表现的最佳实践与常见错误分析示例

评分指标	优良4	可接受3	待改善2	不合格1
合法诚信	关键表现 —— 错误的数量(且) 错误的性质严重程度(或)			
条款齐全				
内容准确	协议内容 无明显错误	服务内容有差错, 或权利义务有差错	收费金额计算错 误,或服务相关时 间计算错误,或服 务内容有差错且 权利义务有差错	HS 编码错误,或 监管方式误用
表达与理解				
谈判技巧				

图8 通过常见错误来描述基于关键表现的不同质量水平示例

评分指标	优良4	可接受3	待改善2	不合格1
合法诚信	关键表现 —— 对客户的违法要求予以严正拒绝 为达成业务而作出不能实现的承诺			
条款齐全				
内容准确	对客户的违法要 求予以严正拒绝; 向客户做出能够 实现的承诺	对客户的违法要求 采取规避策略; 向客户做出条件性 承诺(如配合)	对客户的违法要求采 取模棱两可的态度; 向客户做出较难实 现的承诺(如时间期限)	帮助客户制订违法方 案; 向客户做出完全不能 实现的承诺(如违法)
表达与理解				
谈判技巧				

图9 通过动词、程度副词的变换来描述基于关键表现的不同质量水平示例

析,已经可以确定某一评价指标的关键表现。所谓关键表现,即对该评价指标所对应的任务完成质量产生关键影响的表现。然后以该关键表现为核心,通过动词、程度副词的变换(图9)或者关键表现达成的比例或数量来描述基于关键表现的不同质量水平。

最先确定哪个等级标准,是作为其他等级标准描述的基准。丹奈尔·史蒂文斯建议采用先确定最高等级表现标准的具体描述,再以此为基准逐层描述各低等级表现标准^[21](图9)。特雷西 K·希尔则建议先描述第二高层级(满足成功标准)开始,然后再依次描述最高级(超

出成功标准)和相对较低的层级标准^[22]。在实践中,也可以先描述及格水平的表现标准,再以此为基准分别向两侧逐层描述各高或低等级表现标准。

参考文献:

- [1]冯永琴.职业技术教育学生学业评价的比较及发展[J].中国职业技术教育,2010(12):60-64.
- [2]黄婉圣.高校课堂教学表现性评价存在的问题及解决路径——基于深度学习的视角[J].高教论坛,2020(6):30-32+103.
- [3]赵德成.表现性评价:历史、实践及未来[J].课程·教材·教法,2013,33(2):97-103.
- [4]ASCHBACHER P R.Performance assessment: state activity, interest, and concerns[J].Applied measurement in education,1991,4(4):275-288.
- [5]周文叶.学生表现性评价研究[D].上海:华东师范大学,2009:56.
- [6]GIBBS G.Using assessment to support student's learning[M].Leeds Met Press,2010:24.
- [7]约翰·比格斯,凯瑟琳·唐.卓越的大学教学[M].上海:复旦大学出版社,2015:56-57.
- [8]赵志群.职业教育学习领域课程及课程开发[J].江苏建筑职业技术学院学报,2010,10(2):1-8.
- [9]SPEERS J F.Design and utilization of performance assessment by vocational educators[D].Illinois State University,2008:44-47.

[10]周文叶.表现性评价的理解与实施[J].江苏教育,2019(14):7-11.

[11]ANDRYS O.Constructively aligning the curriculum of a "New Generation" bachelor of environments degree from a social realism perspective[J].Cogent Education,2015,2(1):106-126.

[12]格兰特·威金斯,杰伊·麦克泰格.追求理解的教学设计[M].上海:华东师范大学出版社,2017:176-180.

[13]POPHAM W J.教学课堂教学评价指南:第5版[M].重庆:重庆大学出版社,2010:174.

[14]MARZANO R J,PICKERING D,MCTIGHE J.Assessing student outcomes: performance assessment using the dimensions of learning model[R].Virginia:Association for Supervision and Curriculum Development,1993:29.

[15]殷明.基于大概念与任务分析技术的高职课程学习成果整合设计[J].广东水利电力职业技术学院学报,2020,18(2):71-76.

[16][20][21]丹奈尔·史蒂文斯,安东尼娅·利维.评价量表:快捷有效的教学评价工具[M].广州:华南理工大学出版社,2014:22-28.

[17]HUBA M E,FREED J E.Learner-centered assessment on college campuses: sifting the focus from teaching to learning[M].Needham Heights, MA: Allyn & Bacon,2000:46.

[18]安德森 L W.学习、教学和评估的分类学:布卢姆教育目标分类学修订版[M].上海:华东师范大学出版社,2008:27-28.

[19]彼格斯 J B,科利斯 K F.学习质量评价[M].北京:人民教育出版社,2010:27-32.

[22]特雷西 K.希尔.设计与运用表现性任务:促进学生学习和评估[M].福州:福建教育出版社,2019:105-108.

Design of Task Framework and Rating Scale for Academic Performance Evaluation of Higher Vocational Courses

Yin Ming, Ji Qiong

[Abstract] Higher vocational education focuses on cultivating vocational and application-oriented high-quality technical and skilled talents. Teaching emphasis is placed on ability oriented and work task situations and imparting "functional knowledge". It is more suitable to use performance evaluation to measure students' cognitive and reasoning skills, as well as their ability to apply knowledge to solve practical and meaningful problems. The core of performance evaluation design includes the design of evaluation tasks and rating scales. The evaluation task design framework includes thematic content, actions or practices, context and meanings to effectively stimulate students to demonstrate their acquisition of functional knowledge. The scoring scale design framework includes scoring indicators, indicator weights, quality level, level scores, and standard descriptions, which can serve as a reference for the specific steps of performance-based evaluation design, ultimately promoting better educational effects in higher vocational education curriculum teaching practice.

[Keywords] higher vocational course; academic performance evaluation; task framework design; rating scale design