

# 职业能力测评方法的质量控制

## ——以 COMET 为例

赵志群, 黄方慧

(北京师范大学教育学部, 北京 100875)

**摘要:** 作为职业教育质量评估的重要方式, 职业能力测评已经引起职业教育界的广泛关注, 但是对能力测评方案质量的讨论还显得不足。如果测评结果没有真实全面反映被测者的职业能力发展水平, 即测评质量达不到教育测量技术标准要求, 那么这种测评不但没有价值, 甚至有可能有害。文章讨论了 COMET 职业能力测评的信效度检验方法, 为职业院校在评价、选用和构建科学的职业能力测评方法提供借鉴。

**关键词:** 职业能力测评; 质量控制; COMET

**中图分类号:** G40-011; G71; **文献标识码:** A **文章编号:** 2095-3380(2019)01-0014-05

# Quality Control of Professional Competence Assessment

## —Taking COMET as an Example

Zhao Zhiqun, Huang Fanghui

(Faculty of Education, Beijing Normal University, Beijing 100875)

**Abstract:** As an important way to assess the quality of vocational education, professional competence assessment has attracted widespread attention in the field of vocational education, but the discussion on the quality of the competence assessment is still insufficient. If the results of assessment cannot truly reflect the development level of professional competence, i.e., the quality of the assessment cannot meet the requirements of the educational measurement technical standards, such an assessment is with no value. This paper discusses the methods of reliability and validity test in COMET, which will provide reference for vocational institutions in evaluating, selecting and constructing methods of professional competence assessment.

**Keywords:** Professional competence assessment; Quality control; COMET

近年来,随着“产出导向”教育评估理念的深入人心,职业能力测评(考试)逐渐成为职业教育质量评价研究和实践的一个关注点,全国出现了多种职业能力测评方法和方案,甚至出现了半商业化的能力考试项目。应当说,无论哪种测评方法,只要符合公认的职业教育理念,满足教育测量技术的基本要求,都是有实

践价值的。测量技术的基本要求,是指测评信度、效度和科学性等指标,及由此而“有效构建评价方法、选择标准化测验和合理地解释和使用评价结果”。<sup>[1]</sup>

COMET(competence measurement)职业能力测评方案是对职业院校学生职业能力进行大规模诊断的工具,2017年在教育部督导局组织的高职院校专业评

估预评估中被用作专业评估工具。本文讨论 COMET 职业能力测评信度和效度的检验和保证方法,为职业院校和相关机构在评价、选用和构建科学的职业能力测评方法时提供借鉴。

### 一、COMET 职业能力测评方案简介

科学的职业能力测评需要可信的能力模型和测评模型,模型开发需要相关领域的理论基础。COMET 职业能力测评方案的建立具有成熟的理论基础,这既包括“情境学习”和“多元智能”等公认的普通教育学和心理学理论,也包括“设计导向”“发展性任务”“能力发展逻辑规律”和“工作过程知识”等当代职业教育主流理论。<sup>[2]</sup> COMET 能力模型包含“级别”“内容”和“行动”三个维度,并在此基础上设计测评方案。

其中的级别维度为测评结果提供解释框架,内容维度确定测评内容范围,而行动维度则确定完成任务解决方案的过程。这当中,职业能力级别分为四个等级,即名义能力、功能性能力、过程性能力和整体化的设计能力。按照职业成长逻辑规律,职业学习内容分为四个范围,即职业入门教育、职业关联性教育、职业功能性教育和知识系统化的专业教育。按照行动导向学习理论,完整的行动过程分为六个阶段,即明确任务、制定计划、做出决策、实施、控制和评价反馈。为了开发测试题目,COMET 建立了 8 个能力指标,用以对被测试者的任务解决方案进行解释和评价,这 8 个指标是:直观性/展示;功能性;使用价值导向;经济性;企业生产流程和工作过程导向;社会接受度;环保性;创造性。<sup>[3]</sup>限于篇幅,本文仅讨论 COMET 职业能力测评方案信度和效度的保证和检验方式。

### 二、职业能力测评的信度

信度是指测试结果的一致性 or 稳定性程度,即对同一对象多次施测结果的一致性程度。COMET 能力测评方案在开发和实施过程中采用以下方法考虑和处理信度问题。

#### 1. 评分者信度

COMET 职业能力测评采用开放性测试题目,需要评分者评分,即通过人的判断确定测试结果。评分者的主观判断标准是否一致,是影响测评信度的重要因素。只有对评分者的专业知识、工作经验和个人喜好的影响进行有效的控制,即对评分者间的评分一致性进行控制,才能保证评分者信度,这主要通过“问题

解决空间”和“评分者培训”来实现。

COMET 测试题目来自职业的典型工作任务 (professional task),<sup>[4]</sup> 要求被试人尽可能地利用情境描述所提供的空间,提出任务解决方案并阐述理由。在真实的职业实践中,工作任务的解决不存在绝对的“最优”方案,需要利用现有条件、工具和标准寻找有创意和针对性的解决方案。如果确定了测试题目的标准答案,测试就失去了“职业效度”的价值。为了既不限定答案,又能给评分者提供评分参考依据,COMET 为每一道题目描述了一个“问题解决空间”,确定在所确定的条件下完成任务的可能性,评分者参考解题空间打分。解题空间根据能力水平的 8 项评价指标设计。通常,一个评价指标有 5 个二级指标,解题空间按照二级指标确定和表述(表 1),大约共有 40 个左右。在测评实践中,须按照 Rasch 模型检验一个指标对应的 5 个评分点的同质性。<sup>[5]</sup>如果某个评分点达不到足够的同质性要求,就会被剔除。

表 1 解题空间节选

能力指标	解题空间
创造性	36.答案包含超出问题解决空间的内容
	37.是否提供了一个不寻常、但是有价值的答案
	38.答案具有较高的创新或审美价值
	39.答案显示出了对问题的敏感性
	40.是否充分利用了题目所提供的设计空间

评分者为该专业具有丰富实践经验的教师,在正式评分之前需要接受标准化的培训,以确保理解 COMET 职业能力模型,掌握评分指标和解决空间的含义。评分者培训采用实际测试题目和真实的任务解决方案,通过多次个体评分、小组评分和大组讨论实现一致性。只有当评分者信度符合标准后才能开始正式评分。评分者信度用相关系数 Finn 表示。

严格意义上说,满足这个标准要求等距数据,而且数据必须满足方差分析要求。COMET 能力测评数据是按顺序分级的,但在排列特征为等距、同时对各种特征编号时数字差也是等距的条件下,方可将评分量表作为等距量表使用。方差分析条件包括:观测值的独立性;观测值呈正态分布;方差齐性。违背独立性标准会严重影响分析结果,但违背正态分布或方差齐性标准却影响不大。<sup>[6]</sup> COMET 测评方案满足观测值的独立性要求,这是因为可同时测试一所职业院校所有学生;学生个体的学习方式不一致,观测值也有

一定的独立性;测试中学生独立解答问题,而且平均分配给每个学校每个班级学生不同的测试题目;评分工作在严格控制下独立进行,评分者在评分过程中不交流。在检验数据是否满足正态分布时可使用多种方法,如图形法(直方图、P-P图)和统计法(Shapiro-Wilk检验、Kolmogorov-Smirnov检验)等。

选择单因素评分者间信度测量模型时还要确定,是使用校正过的还是未校正过的评分者间信度作为标准,这取决于评分者的特征。<sup>[7]</sup>由于所有评分者都要对所有被试题目进行评分,而且评分者间信度只针对那些参与抽样的评分者,因此可使用校正过的数值作为评分者间信度标准。只有当被试者间的差别相对较大而对被试者的观测值间方差相对较小时,评分结果才是可信的。Finn系数介于0至1.0之间,在0.5至0.7之间表示及格,大于0.7时为良好。我们认为只有较高的Finn系数才是令人满意的评分者间信度,即控制在大于0.7。在COMET测评实践中,从总体上来看,多数评分者间信度Finn系数都达到了0.8左右。

## 2. 针对测评工具的信度验证

测评工具的信度对整个测评的信度有很大影响。职业能力测评需要考虑工作世界的特殊要求,一些普遍的信度检验方式在此有可能无法使用,或者需要一定程度的变通。

### (1) 再测信度

再测信度是用同一种测评工具对同一组先后两次施测所获得结果的一致性程度,通常用前后测结果的相关系数来表示。如对某一专业同一组学生进行前后两次职业能力测评,如果相关系数越高,则测评信度越高。再测信度受两次施测间隔时间的影响,间隔时间短,试题容易被记住。对于试题数量较少的测试工具来说,会造成假性的高相关;如果间隔时间太长,被试个体的心理特征已经发生变化,导致相关性低。由于职业活动的特点所限,一个职业的典型工作任务一般不会超过20个,COMET职业能力测评试题与一般的数学、外语等考试相比试题数量很少,无法采用通行的再测信度进行信度验证。

### (2) 复本信度

复本信度是用两个复本测验对同一组被试进行先后施测所得结果的一致性程度,要求前后两次测验工具在内容、数量、难度、区分度,时间限制等都要相同,只是表现形式不同,以避免再测信度中个体记忆的影响。由于职业活动的特点所限,我们在正式的

COMET职业能力测评中不采用复本信度进行信度验证,但是在预测试中采用它保证题目的效度。

### (3) 折半信度

折半信度是指在没有复本的情况下,将测验试题分为对等两半,被试在两半测验题目中所得结果的一致性程度。折半信度与复本信度相似,但没有将测试分为两次进行,只是在施测后人为将试题分半。在COMET职业能力测评中,测试题目是综合性的开放性试题,每个题目反映的都是一个职业实践共同体(practice community)的典型工作情境,不可能把试题分半,因此无法进行折半信度的检验。

### (4) 内部一致性信度

科学的测评工具必须保证所有测试试题都指向同一个概念,即保证内部的一致性信度(也称同质性信度),反映测验内部所有题目间的一致性程度。<sup>[8]</sup>如果职业能力测评的同质性信度高,说明测验内部条目的一致性和稳定性高,测验测的是同一种职业素养;如果同质性低,说明测评结果可能是几种素养的综合反映。测试内容的同质性是信度要求的重要组成部分。COMET测评通过理论构想对个体的心理特征的进行诊断,只有同质性信度高,才能根据分数对职业能力发展状况做出合理的推论和解释。在COMET测试题目开发过程中,我们特别关注对异质性题目的剔除。王荣秀以高等职业教育建筑工程专业为例,对测试题目的开发过程和剔除办法进行了系统性的研究和验证。<sup>[9]</sup>

## 三、职业能力测评的效度

效度是指测量结果的有效性或正确性程度。在职业能力测评中,它是测评工具能够测到的职业能力的程度,效度越高,越能反映职业能力发展的真实情况。从心理测评技术角度看,在选用或开发职业能力测评工具时,必须对其效度进行评估。COMET测评方案从以下方面保证其效度。

### 1. 内容效度

内容效度是指实际评价的内容、范围与所要评价的内容、范围相符合的程度,它体现测评内容在多大程度上能代表受测的内容范围,即测评内容是否能够反映测评目标。<sup>[8]</sup>在通常的教学实践中,教师通过各种测验对学习成果进行监控,测试内容多为课堂所学的重要知识。COMET职业能力测评的目标是诊断学生的职业能力发展情况,测试内容应反映学生从事一

门职业所需要的能力,即反映“职业的效度”。由于不同院校课程设置不完全相同,采用的具体学习任务差别更大,不可能通过对某一门或几门课程内容的测试了解学生的整体职业能力发展水平。按照情境学习和能力发展逻辑理论,每个人“从初学者到专家”的成长过程中所要完成的发展性任务的内容是类似的,<sup>[10]</sup>职业能力测评的内容是职业工作,而不仅仅是在某一课程中所学的知识和技能点。按照职业工作的效度开展测评,使不同地区、甚至不同国家间职业教育质量的比较成为可能。

在职业教育中,当测试题目符合真实的职业任务时,便认为其具有内容效度。为了保证内容效度,避免试题设计者的主观经验影响内容效度,COMET 测试试题开发采用实践专家研讨会的方式,试题反映职业实践的典型职业工作情境。测试题目的内容效度一般通过专家评定法进行检验。由于专家评定法的主观性较强,采用克伦巴赫统计弥补其不足,即在试题预测阶段进行复本间相关系数计算,以确保内容效度。来自本行的专家(教师)在接受评分者培训后,可以评估测试题目对被测试群体的内容效度。由于测试题目的内容效度与真实职业情境有关,而不是与课程相关,因此不同层次职业院校的相同专业的学生可参加同一次测评。如中职、高职和应用型本科的护理专业学生共同参加测试,结果表明,大家均认为这些测试题目具有内容效度,因为作为未来医院不同层级的护士,他们都需要完成护士的典型工作任务。这些学生普遍需要具备护士的执业资格证书。

## 2. 构念效度

当测试结果以准确且可理解的方式反映被测量的构念(如职业能力)的内涵时,便具备了构念效度。<sup>[11]</sup>构念效度的检验需要特定的理论基础,通过理论引导出各项关于潜在特质的基本假设,用实证方法核查测量结果是否符合理论假设。在职业能力测评中,构念效度具有特殊的意义,因为我们很难为内容效度确定一个客观的数值,职业教育的能力测评也无法确定适当的外部指标。我们只能从目标构念(即COMET能力模型)中推导出假设,然后再依据对假设的检验计算出构念效度。COMET研究团队通过对能力模型进行心理测量评估,使之成为一个测评模型,<sup>[12]</sup>并通过结构方程模型中的验证性因素分析,检验其构念效度。<sup>[13]</sup>

## 3. 效标关联效度

效标关联效度是指测试结果与一个外部效标的一致性程度,即测试结果能够在多大程度上代表或预测效标行为。<sup>[8]</sup>效标关联效度分为同时效度和预测效度:同时效度指在同一时间内获得的测试结果与效标行为之间的一致性程度;预测效度是指测试结果对效标行为的预测程度。效标关联效度检验的关键是效标的选择,效标必须是为社会所接受的、能够具体反映某项特定内涵的指标。<sup>[11]</sup>

### (1) 同时效度

效标关联效度的评估通过对测试结果与效标的相关性分析进行验证。对于职业教育而言,可以参照通行的职业资格考试或者技能大赛确定效标效度。例如在护理专业学生职业能力测评中,将被试学生护士执业资格考试成绩作为效标,发现测试成绩与护士执业资格考试成绩显著相关,说明职业能力测评方案具有较好的效标关联效度。<sup>[14]</sup>针对内容效度不足的问题,我们还采用了相反的做法,即按照COMET能力测评方法对现有职业技能大赛进行评估。只有当技能大赛同样也以企业真实工作任务方式测试职业能力(技能)的全部方面时,才有可能将此类考试作为外部的效标效度指标。王珺燕以职业院校技能大赛嵌入式产品开发竞赛项目为例,以COMET职业能力模型作为参照标准,通过吻合度研究分析竞赛试题内容与职业能力要求的匹配情况,阐释了技能大赛试题的职业效度。研究发现,技能大赛试题和评分标准主要反映选手“做了什么”,部分注意到了“怎样做的”,但很少关注“为什么而做”,因为学生的行动可能只是习惯与固化的结果。<sup>[15]</sup>这恰恰证明了COMET能力测评在过程性能力与设计能力级别的职业能力测评的效度。

### (2) 预测效度

职业能力测评在学术界也受到一些质疑,主要疑问是:COMET采用纸笔测试方式评价学习者的认知能力,认知能力测评结果在多大程度上能反映被测者在真实工作情境中的行动表现水平?这就是所谓的职业能力测评的预测效度问题。

周瑛仪等对能力测评对受测者心理特质或未来绩效所作预测的准确度问题进行了深入的探讨。职业能力可以分为认知部分、通过行动表现出来的部分以及这两者之外的其它部分。作为操作性定义,她将职业能力中的认知部分称为职业认知能力,将通过行动表现出来的部分称为职业行动能力。按照“以一定的理论假设为指导,能反映理论假设和研究变量的主要

维度,而且简明可行”<sup>[16]</sup>的原则选择测量指标,采用“明确任务”和“制定计划”两个指标测量职业认知能力,采用“现场分析”“准备维修清理”和“评价”测量职业行动能力。选择汽车维修、数控加工和电气自动化三个专业,以职业典型工作任务为基础开发测评任务,要求被试者完成 COMET 能力测评,以评价他们的职业认知能力;之后要通过实操解决问题,以评价他们的职业行动能力,采用相关分析和验证性因素分析进行预测效度检验。结果发现,COMET 职业认知能力测评的结果能够预测职业行动能力,大规模职业能力测评能够对职业能力进行比较全面、有效的评价。但是在不同专业领域,职业认知能力对职业行动的影响机制存在差异。若在现有测评内容基础上,增加对现场分析过程中认知能力的评价,将会进一步提高职业认知能力测评的预测效度。<sup>[17]</sup>

#### 四、结语

能力是个体的心理特征,一般认为是不可能被准确测量的。在心理测验领域,不能直接测量或观察的抽象构念,一般通过可测量的变量进行间接测量,如利用问卷、量表或行为测量等工具。<sup>[11]</sup>目前职业能力测评的方法很多,但是很多方法没有对信度和效度进行足够的考量,而这两项指标恰恰是保证测评工具质量的核心要素,尤其是采用定量测评的时候。就职业能力测评而言,效度比信度更为重要,因为首先必须保证测评工具能够真实测出职业能力的水平。否则,信度再高,测评结果也没有实际意义。在对职业能力测评工具的质量进行评价时,必须同时考虑信度和效度标准。具有较高信度和效度的测量工具才能测出有价值的测评结果。由于篇幅所限,本文没有罗列详细的检验过程。但是基于本文所讨论的检验方法,可以对职业能力测评工具做出科学的评价,促使职业教育质量评价工作更加严谨。

#### 参考文献

- [1] [美]诺尔曼.E.格朗伦德.学业成就测评(第七版)[M].南京:江苏教育出版社,2008:160.
- [2] 赵志群,(德)劳耐尔.COMET 职业能力测评方法手册[M].北京:高等教育出版社,2018:26-52.
- [3] Rauner F, Heinemann L, Maurer A, et al. Competence Development in TVET (COMET),Theoretical Framework and Empirical Results [M]. Dordrecht:Springer, 2013:3-4.
- [4] 赵志群.职业教育工学结合一体化课程开发指南[M].北京:清华大学出版社,2009:33.
- [5] 何兴国,赵志群,何壮.德国 COMET 评分量表在我国应用的多面 Rasch 模型分析[J].职业技术教育,2016, 37(25): 69-74.
- [6] Bortz J, Dring N. Forschungsmethoden und Evaluation [M]. Berlin: Springer,2002:180.
- [7] Shrout P E, Fleiss J L. Intraclass Correlations: Uses in Assessing Rater Reliability [J]. Psychological Bulletin, 1979,86 (2):420-428.
- [8] 张洪秀.教育测量与评价方法[M].长春:吉林大学出版社,2014:53-54,61,67.
- [9] 王荣秀.基于 COMET 的职业能力测评试题开发研究——以高等职业教育建筑工程专业为例[D].北京:北京师范大学,2015.
- [10] Rauner, F. Berufliche Kompetenzentwicklung von Novizen zum Experten[A]. // Dehnbostel, P. (Ed.). Kompetenzentwicklung in vernetzten Lernstrukturen [M]. Berlin: Edition sigma, 2002.
- [11] 邱皓政.量化研究与统计分析——SPSS(PASW)数据分析范例解析 [M]. 重庆:重庆大学出版社,2013: 304-305, 294.
- [12] Martens Th, Rost J. Zum Zusammenhang von Struktur und Modellierung beruflicher Kompetenzen [A] . // Rauner F, Haasler, B, Heinemann L Grollmann, P. (Eds.). Messen beruflicher Kompetenzen. Bd. I. [M] Münster: LIT,2009:91-95.
- [13] Martens Th, ua. Ergebnisse zum Messverfahren COMET [A]. // Rauner, F. Messen beruflicher Kompetenzen. Bd. III. Drei Jahre COMET -Testerfahrung [M]. Münster: LIT,2011:90-126.
- [14] 田甜, 庄榕霞.KOMET 护理专业职业能力测评方案的有效性研究[J].职教论坛,2016(15):32-37.
- [15] 王珺燕. 嵌入式产品开发技能竞赛试题与职业能力模型的吻合度研究[A].庄榕霞,赵志群.职业院校学生职业能力测评的实证研究[M].北京:清华大学出版社, 2012:164-191.
- [16] 董奇.心理与教育研究方法(第三版)[M].北京:北京师范大学出版社,2008:110,125-126.
- [17] 周瑛仪.大规模职业能力测评的预测效度——基于 COMET 方案在汽修、数控与电气专业领域的研究[D].北京:北京师范大学,2015.