

如何基于标准命题： 从双向细目表走向测验设计框架^{*}

文 崔允漷 邵朝友

教师基于标准的命题是实施基于标准评价的前提,也是教师常态工作之一。作为命题的一种重要工具,双向细目表并没有很好地体现出基于标准命题的要求。本文在分析双向细目表存在问题的基础上,并加以改进,提出一种测验设计框架,从而为教师基于标准的命题提供一种参照。

基于标准的命题是指教师以课程标准为依据,进行测试题目的开发,并使得测验与课程标准相一致。要实施基于标准的评价,仅仅依靠高风险测试(如高考、中考)的引导是远远不够的,课堂层面上教师的测验是主阵地,也是最终的落实点。而命题是测验的准备,这必然引申出一个重要的话题——教师如何基于标准命题。遗憾的是,在实践中,教师往往凭经验编制试卷,甚至随意挑选题目组合试卷。事实上,教师选用的测试题题目也大多不是评价专门机构或专业人员开发、编制的。无疑,如何有效地为教师提供一种基于标准命题的框架就成为当下之急。

一、双向细目表在基于标准命题中存在的问题
命题需要一定的分析工具和设计方案,双向细目表就是其中最常用的一种。双向细目表由考察内容和考察目标组成的列联表,其常见的格式与内容如表1所示。一般而言,双向细目表包含三个要素:考察内容,如课程标准中规定某个单元知识;考察目标,如课程标准规定的某个知识点的认知要求;考察内容和考察目标的比例/权重。

表1 双向细目表的常见格式

考察内容	考察目标			
	学期/单元	记忆	理解	应用
知识点1		_____%	_____%	_____%
知识点2				
.....				

不可否认,从操作的角度看,双向细目表科学、规范地为教师命题提供了便利。如从内容及认知要求上,双向细目表具有较高的内容效度,也比较好地回答了一些该门考试在能力上“考什么”的问题,体现了基于标准命题的一些特点。但是,双向细目表也存在不足。

首先,一个明显的缺陷是考察目标范围太窄。在中国学术期刊网上以“双向细目表”为关键词进行搜索,可查到151篇文章。这些资料基本上把双向细目表中的考察目标简单定为:记忆、理解、应用,并没有全面反映课程标准规定的认知要求。

其次,双向细目表关注于测试的内容,而命题的其它必要因素,如试题背后所蕴涵的价值问题、评分细则、难度、题型、题量等,并没有得到应有的重视。

最后,双向细目表只是命题的一个组成部分,缺乏总体规划。命题是由一定的程序组成,如命题目的、题目的评分方法、测试时间等,但双向细目表只关注了试题编制中的一部分。

二、测验设计框架:基于标准命题的蓝图

为了更好地促进教师基于标准进行命题,在吸取、保留双向细目表优点的基础上,针对上述问题,我们对双向细目表做进一步的改进。

1. 双向细目表的一次修正

命题作为评价的重要环节,自然需要与课程标准保持一致,从相关的文献来看,评价与课程标准的

* 本文是教育部人文社会科学重点研究基地重大项目《基于课程标准的学生学业成就评价研究》的研究成果之一。

一致性可从 7 个纬度来考察 见表 2)

表 2 评价与课程标准一致性的四种分析框架^②

纬度模型	测试内容	认知要求	范围	难度	题量	教学引导	价值取向
韦伯 (Norman L. Webb)	评价内容与课程标准内容是一样的	评价要求的认知深度与课程期望学生达到的要求一样	课程标准期望学生学习的知识范围与评价需要学生正确回答所需知识范围一样	所考察的课程目标获得均衡的测试题目数量			
成就公司 (Achieve, Inc.)	该标准用来检查评价内容与课程标准内容是否一致的成一致的	该标准关注评价要求的认知深度与课程期望学生达到的要求一致的程度	设置该标准的目的是为了判断测试题目中是否含有具有挑战性的学科问题	每个课程标准内容有相应的题目来评价:一个内容的不同认知要求都被评价			
威克森 (Karen L. Wixson)	课程标准的内容均衡地分布于测试的题目中	评价的认知要求与课程标准要求一致	评价范围与课程标准范围一致	所考察的课程目标获得均衡的测试题目数量			测试题目隐藏的价值观与课程标准要求的价值观一致
2061 计划 (Project 2061)	测试题目必须反映出相应的课程标准目标	测试题目还需要充分地反映出相应的课程标准的具体目标	评价应关注课程标准内容的“理解”层次,“理解”是高于记忆的高级认知水平,评价包含学生熟悉和新奇任务	教师理解学生的反应,教师能从评价材料中获得具体的建议,并能据此策出下一步行动所要解决的问题			

其中的测试内容、认知要求对应于双向细目表中的考察内容和考察目标。对于认知要求,我们可借鉴新修订的布鲁姆目标分类。作为对前人的超越,新修订的布鲁姆目标分类非常完整,被认为是认知目标分类的权威。从表 3 中,我们可以看出新修订的布鲁姆目标分类中的认知过程维很好地概括了学生学习活动的表现方式。

表 3 新修订的布鲁姆认知目标分类中的认知过程维^③

水平层级	内容
记忆	包括识别、回忆;
理解	包括解释、举例、分类、总结、推断、比较、说明;
应用	包括执行、实施;
分析	包括区分、组织、归属;
评价	包括核查、判断;
创造	包括生成、计划、贯彻。

仔细研究我国各门学科的课程标准,从中可以发现新修订的布鲁姆目标分类很好地囊括它们所规定的认知要求。因此,我们认为有必要在运用双向细目表时,要拓宽其考察目标,即把考察目标记忆、理解、应用修订为记忆、理解、应用、分析、评价、创造。这样,从认知要求的角度确保了命题与课程标准的一致性。

范围则主要指测试范围与标准范围一致,从实际情况而言,教师很少在此出错,因此,我们并没有引进“范围”考量。教学引导表明的是大规模考试对所有教

师教学的引导作用,而教师的命题主要为自己日常教学服务,针对任教班级学生或年段,其辐射作用对象主要为同教研组。更重要的是,高质量的试卷自然也就具有教学引导作用,因此我们也忽略了该纬度。

题量则表示测试题目均衡地分布于各测试所考察的课程目标。由于大规模考试,特别是中考、高考,是高利害的,对教师课堂层面的命题具有导向作用。考试的重点,也就是常说的热点,往往成为教师命题的重心,那些相对“冷门”的课程内容则往往得不到重视。事实上,这些“冷门”内容也是课程标准规定学生必须掌握的课程内容。我们倡导以课程标准为依据进行教学,平时教什么就考什么,因此,我们认为有必要把题量作为命题的一个考量。这有助于纠正那种考什么就教什么的错误倾向,也有助于教师开展基于标准的教学。

命题并非价值中立,它还具有一定的价值取向,应与课程标准所追求的价值观保持一致。令人遗憾的是,实践中很多命题者以为测试只是一个技术问题,以至编制出来的一些测试题目让人哭笑不得。如有这样的一个测试题目:如果一颗子弹以 100 米/秒射穿一个敌人的头后,其速度变为 50 米/秒,请问子弹大约可以共打穿几个敌人的头。显然,题目所体现出的“阶级情结”和暴力倾向与课程标准所追求的和平的主题是相违背的。高质量的题目不仅是手段,同时也是目的。这是因为在应答过程中,学生还将感受到题目内在的意义。

最后,我们保留难度这个考量,这是因为标准虽然是底线,但教育拒绝平均化,追求卓越,基于标准的评价关注的是高水平而又可以达到的标准。

2. 双向细目表的二次修正

除了上述考量外,一般而言,常用的命题考量还应包含命题的目的、题型、评分细则、测试时间。之所以如此,其理由如下。

命题是项针对性很强的活动,起始于目的,目的是命题的灵魂。这些目的包括,考察学生在预期的学业水平上所达到的程度、如何采取措施改善学生的学习等。而不同的目的直接影响试卷的难度,如地区统考中的高考模拟测试难度就可能比会考模拟测试难度大。可以说,命题目的的引导作用贯穿于命题的整个过程。

题型则要求教师确定试题的类型,如选择、匹

配、是非、论述题等,也就是选择评价方法的问题。评价方法合适与否自然也就影响了评价的效度。由于不同题型适用的对象不同,教师必须依据评价目的与内容做出正确抉择。

同时,不同的题型要确定相应的评分方法,如主观题的评分方法需要相关细则,严格要求时还需要制定评分规则。评分细则的质量反映了教师对课程标准和学生表现的把握程度,它直接影响了评价结果的信度。

此外,测试所需时间也是一个重要的纬度。一般来说,试卷拼出后,由教师做预答时,其答题时间应以学生考试规定的时间的3/5或1/2为宜。

3. 测验设计框架的形成

综上所述,改进后的双向细目表呈现如下(见表4),为了便于陈述,我们称其为测验设计框架。由于课程和教师经验层次的差异,需要指出的是本表只是提供一种命题分析框架,命题时,不同的教师应视具体情况灵活加以应用。

表4 测验设计框架

命题的目的与试卷的总体难度:												
考察内容	考察目标							其他考量				
	学期/单元	记忆	理解	应用	分析	评价	创造	难度	题型	题量	价值取向	评分细则
知识点1												
知识点2												
.....												
测试时间:	分钟											

三、测验设计框架的特点与优势

1. 测验设计框架的特点

内容上全面地体现了基于标准命题的要求。在常态的教师命题情况下,测验设计框架所包含的内容(如考察内容与考察目标、题型与题量、难度与价值取向、评分细则)与课程标准(如知识及认知要求、难度与课程的价值观、考察知识的题目数量之间的平衡)达成了深度匹配。

结构上显示了基于标准命题的一般程序。命题程序具有一定的规范要求,如Robert L. Linn & Norman E. Gronlund和加州旧金山市的西部教育研究实验室(WestEd)提供的基于标准的测验开发步骤就具有一定的代表性。相比而言,测验设计框架提供的程序更为全面:先确定明确测试的目的,进而明确试卷的总体难度;确定考察内容与考察目标;明

确测试题目的难度,并选择合适的方法或题型;考察每道题目隐藏的价值取向;最后再确定测试时间和制定各个题目的评分细则。

2. 测验设计框架的优势

相对双向细目表,测验设计框架在命题实践中具有独特的优势,这些优势包括:

规范了教师基于标准的命题。测验设计框架以课程标准为依据,全面地反映了课程标准的内容与要求,也体现出命题的一般程序,从而为教师基于标准命题提供了一种分析框架,在一定程度上消解了命题的随意性与盲目性。

促进了基于标准评价的落实。当教师把测验设计框架作为命题规范之时,就是基于标准命题之刻。这也为课堂层面上大规模落实基于标准的评价提供了可能,也极大地促进了评价与课程标准的一致性。而追求评价与课程标准的一致性恰恰就是基于标准命题的意旨所在。

提升了教师的评估素养。命题是项综合性很强的技术,涉及了很多因素,如已有题目的选择、题目类型的确定、各类题目权重分配等。正因为命题包含众多的因素和技术,教师只有真正进行科学、规范的命题,才能掌握其中各种技术。同时,高质量的命题能对其他命题者产生积极的影响。当一份好试卷被其他命题者共享后,他们能从中反思自身命题中的缺陷与不足,并为他们改进命题提供了一种可能。

四、预计的问题与建议

直面我国当前的教育现状,我们可清醒地预计到测验设计框架在实践中将不可避免地际遇各种不利因素。这些不利因素在教师、学校、教育政策层面上尤其得以凸显。

教师层面:应用测验设计框架的不利因素积重已久。阻碍测验设计框架应用的主要因素有——专业活动时间的杂乱、片段化无法提供应用测验设计框架所需的充分时间,或者说教师缺乏充分的课程资源——时间;教师对学生学习评估的专业技能之不足;教师的专业活动很大程度上受到本校、本教研组的“习惯做法”的影响,很难打破自身原由的思维习惯。

学校层面:基于标准命题的各项保障措施被严重忽视。这些保障措施如,在学校课程规划与实施

中,如何建立学生学业评价制度,如何为促进教师基于标准命题而制订教师评价、考核、奖励制度等。其实,以上问题反映了学校对教育的理解和诠释,也直接关系到测验设计框架被采用与否。

教育政策层面:基于标准命题的配套措施还不成熟。虽然课程改革已历经一段时间,但关于如何促进教师基于标准命题的政策还有待出台。如怎样建立问责制度、设立问责机构,如何在教师教育或教师专业标准中增加教师评估素养培养要求,如何进一步完善高风险测试。这些措施的缺位无疑直接影响到教师、学校对测验设计框架的关注程度。

如何解冻这些固有的状态,实现从旧状态到新状态的变革?虽然解决上述问题的方法并没有一个简单明确的答案,但我们认为,至少可从三个方面做出决策。

教师应努力提高自身评估素养,积极尝试应用测验设计框架。对于教师而言,首先要多方谋求提升自身评估素养的途径,认识到评价对于课程实施、学生学习的作用,并不断提高评价技能。同时,在命题时能积极应用测验设计框架,使之从文本转化为行动。

学校应创设各种条件,鼓励教师积极应用测验设计框架。学校应充分认识到优质评价对于教学质量的意义,要有计划地开展评价理论的学习。课程实施过程,有意识地向教师推荐测验设计框架,鼓励教师积极进行尝试,并把测验设计框架作为教师命题的一种规范,必要时以此作为一种教师考核、评价的指标。

相关教育专业部门应加强评价与课程标准的一致性研究。应用测验设计框架的实质是如何使得评价与课程标准保持深度匹配。在我国,评价与课程标准一致性的研究基本还处于真空状态。因此,对评价

与课程标准一致性的研究与上述的基于标准命题的配套措施的制订与实施,应同步进行,从而为测验设计框架的应用创设良好的条件和氛围。

总之,在教育问责加大的情况下,所有的教育部门及相关人员都应充分认识到基于标准命题的意义,自觉地投入到基于标准命题的研究与行动中,把测验设计框架作为教师命题的一种常态工具。

注释:

截止日期为 2007-3-29, 搜索时间范围为 1994-2006 年。

Rothman R. (2003). Imperfect Matches: The Alignment of Standards and Tests.

<http://www7.nationalacademies.org/bota/Alignment%20Standards.doc>.

有多种模型分析评价与标准的一致性,由于这些模型有很多内容是重叠的,在此仅列举四种典型框架。

盛群力等. 布卢姆认知目标分类修订的二维框架[J]. 课程·教材·教法, 2004,(9): 90~93.

评分规则(rubric) 是一种学生学习评价工具,依据课程标准的要求,它详细地描述出学生在不同学业水平上的表现。

张增学,高艳华. 考试命题中的几个问题[J]. 中等医学教育杂志, 1994,(2): 49~50.

[美]Robert L. Linn & Norman E.Gronlund 著. "促进教师发展与学生成长的评价研究"项目组译. 教学中的测验与评价. 北京:中国轻工业出版社, 2003. 该测验开发步骤为: 确定测量的目的; 设计细目表; 选择合适的测验题目和评价类型; 编制一套适宜的题目和任务。这里的细目表指的是测验双向细目表。

Kim O'Neill, Kendyll Stansbury (2000). Developing a Standards-Based Assessment System: A Handbook. WestEd, San Francisco, CA. 该测验方案包括: 明确评价内容和认知行为类型、编制试题(题型), 有时还要列出试题的难易度。

王建军著. 课程改革与教师专业发展[M]. 成都: 四川教育出版社, 2004.

[崔允漷 邵朝友 华东师范大学课程与教学研究所 200062]

长三角信息平台

上海市崇明县举行第九届教育科研暨第三届教师科研成果颁奖大会

日前,上海市崇明县教育局召开了崇明县第九届教育科研暨第三届教师教育科研成果颁奖大会。崇明县教育局领导和各校分管科研的校、园长、教科室主任、获奖者代表共 250 多人出席了本次会议。

崇明县教育局副局长黄强出席大会并作了“让教育科研成为教师专业发展的原动力”的讲话。他既对本届科研成果的特点和存在问题进行了全面而系统的归纳与分析,又对新时期我县教育科研工作从县科研部教科室、学校校长、学校科研室三个层面提出了新的要求和希望。他的讲话为打造崇明生态教育的品牌和建设具有新郊区新农村特色的教育科研事业指明了方向。

(刘伟超)