

区域基础教育学业评价与课程标准一致性的本土化研究

——以北京市为例

田 一 胡 玲 李英杰

【摘要】区域基础教育学业评价是地方教育行政部门监测基础教育质量的重要方法，而国家课程标准是基础教育学业评价的根本和直接依据，因此，分析区域基础教育学业评价与课程标准的一致性对于开展基础教育评价改革可谓至关重要。本研究借鉴国际上学业评价与课程标准一致性分析的经验，结合国内教育教学实际情况，开发了本土化的一致性分析模式，并对北京市17个区域小学语文、数学、英语、科学、品德与社会等学科学业评价工具进行了一致性分析。结果表明，区域基础教育评价工具的导向性偏低，与课程标准内容目标的匹配程度较高，与能力目标的匹配程度较低。针对这些问题，学业评价可在研发学科学业标准、规范命题流程环节、加强命题专业化培训、增加实践类评价手段上做进一步努力。

【关键词】区域评价；课程标准；一致性；本土化

【中图分类号】G527 【文献标识码】A 【DOI编码】10.16518/j.cnki.emae.2016.10.006

《国家中长期教育改革和发展规划纲要（2010—2020年）》提出，要“提高义务教育质量，建立国家义务教育质量基本标准和监测制度，严格执行义务教育国家课程标准”。从这个角度看，国家课程标准是义务教育质量监测和基础教育学业成就评价的根本依据。目前，各省市纷纷开发各种学业评价工具开展基础教育质量监测工作，进行区域基础教育学业评价体系研究。可是，已有的区域基础教育学业评价是否完全落实了课程标准？它们与课程标准是否真的具有一致性？

目前，国际上关于一致性分析的研究主要集中在美国。20世纪80年代，美国发起了“由标准驱动并基于标准”的基础教育课程改革，许多研究者和研究机构提出了学业评价与课程标准一致性分析模式。其中，韦伯模式、Achieve模式和SEC模式最具代表性。韦伯模式是美国学业成就评价领域知名学者诺曼·韦伯（Norman L. Webb）于1997年提出的，主要以内容重点为核心，从内容领域一致性、知识深度一致性、知识广度一致

性、分布平衡性四个维度测查学业成就评价与课程标准内容要素的匹配程度。^[1]而Achieve模式以罗斯曼（R. Rothman）的理论为基础，由美国非赢利教育研究机构——成就公司（Achieve, Inc）组织开发了一致性研究的综合化工具，主要从向心性、挑战性、均衡性三个方面构建综合性较强的新型“学业评价—课程标准”一致性研究工具。它具有浓重的韦伯工具的印迹，内涵与之接近。^{[2][3]}SEC模式则是由威斯康星州教育研究中心学者A. Porter和J. Smithson等人以计划课程调查（survey of enacted curriculum，简称SEC）数据为基础，共同开发和研制出评价与课程标准一致性水平的分析程序和方法，它强调一致性研究的整体化，主要用于进行课堂教学与学业评价之间的一致性比较。^[4]此外，威克森（Karen K. Wixson）模型、TIMSS测验—课程匹配分析、斯坦福国际咨询研究所（Stanford Research Institute International，简称SRI International）的模式、“2061计划”（Project 2061）等也各有侧重^[5]，从不同方面丰富和深化了一致性的内涵，有效弥补了各研究工具实施的

本文为北京市教育科学“十二五”规划青年专项课题“基础教育学业评价与课程标准一致性分析模式的开发与应用”（项目编号：CBA14047）的阶段性研究成果。

田 一 /北京教育科学研究院助理研究员，博士，研究方向为教育测量与评价。（北京 100191）

胡 玲 /北京教育科学研究院高级教师，研究方向为品德与社会课程与评价。

李英杰 /北京教育科学研究院高级教师，博士，研究方向为语文课程与评价。

不足。

在国内,华东师范大学崔允漷团队对基于标准的学生学业评价的一系列研究,引起了学者对学业评价与课程标准一致性研究的关注。^[6]当前研究主要集中在两个方面:一是美国的一致性分析模式介绍^{[7][8][9][10]};二是基于美国模式的学科应用研究,既包括学科测评工具与课程标准的一致性分析,也包括学科教材与课程标准的一致性分析,前一个方面的研究反映出学科测评工具与课程标准未能完全保持一致,存在目标领域下知识掌握水平界定模糊、无具体学业标准可参照、教育命题技术欠缺等问题^{[11][12][13]},后一个方面的研究则表明学科教材在知识点分布、认知水平要求上与课程标准具有较好的一致性^[14]。由此可见,已有研究成果主要集中于针对某单一区域、单一学科而开展的一致性分析研究,而大规模针对不同区域各学科学业评价与课程标准一致性分析的实证研究还比较缺乏。因此,立足宏观视角,进行开发区域学业评价与课程标准一致性分析的本土化工具的相关实证研究显得尤为迫切。

综上所述,依据国内基础教育改革和发展的相关政策规划,充分借鉴国内外一致性分析研究经验,结合我国实际教育教学评价情况,开发本土化一致性分析工具,具有理论与实践的双重意义。一方面,它弥补了相关实证研究领域的空缺;另一方面,它有助于我们更加科学有效地分析区域基础教育学业评价与课程标准的一致性情况,推动基础教育课程改革有效、顺利地开展。

一、研究方法与过程

(一)研究对象

北京作为中国的首都,在教育政策落实和评价改革方面均占据领先地位。本文的研究对象为北京市17个区域小学阶段监测学科评价工具(用于期中或期末考试、区级学业质量监测等),共147套,其中包括44套语文学科评价工具,44套数学学科评价工具,43套英语学科评价工具,9套科学学科评价工具,7套品德与社会学科评价工具。这些工具均包括试题和评分标准,部分工

具含有框架、细目蓝图和试题分析等内容。

(二)研究工具

借鉴前文介绍的韦伯模式、Achieve模式和SEC模式的分析框架,依托相关国内外研究与已有实践,笔者构建了本研究的分析框架,如图1所示。

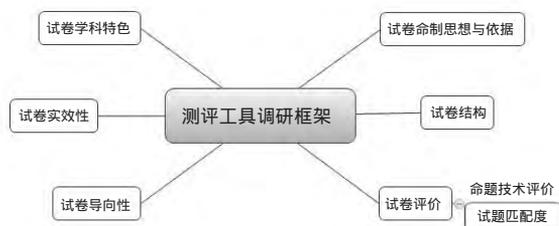


图1 区域基础教育学业评价与课程标准一致性分析之框架

基于研究框架编制的研究工具包括试卷内容评定问卷和试题内容评定表。其中,前者从试卷的整体角度,对命题思想与依据(7道题目)、试卷结构(7道题目)、试题评价(11道题目)、试卷导向性(7道题目)、试卷实效性(2道题目)、试卷学科特色(5—9道题目)等方面进行调查,采用李克特量表的形式,从“完全不符合”到“完全符合”进行五点计分;后者从试题的角度,逐题评定其与课程标准内容目标、能力目标的匹配程度,与细目蓝图内容目标、能力目标的匹配程度,从“完全不匹配”到“完全匹配”进行五点计分。

经教育测量学指标分析可知,该研究工具的内部一致性信度达到0.97,各维度与总分之间的相关系数均在0.91—0.93之间,信、效度较好。

(三)研究方法与设计

1. 专家评价判断法

本研究采用专家评价判断法,聘请教育评价领域专家与学科教学评价专家,依据国家课程标准等相关文件要求及命题测试与调查工具的技术要求,在结合区域教材版本及区域教育教学实际情况的基础上,对区域评价材料进行逐卷、逐题评价。各学科领域的评价专家人数分布如下:小学语文18名,小学数学18名,小学英语18名,小学科学5名,小学品德与社会4名。

2. 双盲实验设计

本研究采用双盲实验设计,即在研究准备阶

区域基础教育学业评价与课程标准一致性的本土化研究

段对各区域材料进行匿名处理,研究过程中,专家不知道自己所评价的材料属于哪个区域,各区域学业质量监测人员也不知道本区的材料被哪位专家评价。

(四)分析方法

本研究综合了定量与定性的分析方法,其中,定量分析是通过 SPSS20.0 统计分析软件对专家评定量表的数据进行描述统计和频次分析;定性分析是基于各区域评价工具和专家建议进行文本分析。

二、研究结果及讨论

(一)总体结果

表 1 总体试卷和试题评定结果表

	评定频次	最小值	最大值	平均分	标准差
加权总分	603	2.02	5.00	3.97	0.62
试卷命题思想依据	603	1.75	5.00	3.92	0.69
试卷结构	603	1.67	5.00	3.88	0.64
试题评价	603	2.42	5.00	4.14	0.45
试卷导向性	603	1.29	5.00	3.77	0.73
试卷实效性	603	1.00	5.00	4.08	0.79
试卷学科特色	603	1.50	5.00	3.94	0.72
与课程标准内容目标的匹配程度	603	1.65	5.00	4.64	0.52
与课程标准能力目标的匹配程度	603	1.29	5.00	4.47	0.69
与细目蓝图内容目标的匹配程度	216	1.83	5.00	4.77	0.41
与细目蓝图能力目标的匹配程度	174	2.43	5.00	4.66	0.40

注:1.加权总分=命题思想依据 $\times\lambda_1$ +试卷结构 $\times\lambda_2$ +试题评价 $\times\lambda_3$ +试卷导向性 $\times\lambda_4$ +试卷实效性 $\times\lambda_5$ +试卷学科特色 $\times\lambda_6$ (其中 λ_i 表示各维度的权重)。

2.题目匹配度是根据《题目内容评定表》的各题目评分加和平均计算而得。1—不匹配 2—较差匹配 3—一般匹配 4—较好匹配 5—完全匹配。

从表 1 中总体试卷的评定结果来看,区域总体试卷平均分为 3.97 分,整体处于一般偏上水平。试题命题指导思想依据、试卷结构、试题评价、试卷实效性、试卷学科特色各维度的平均分均在 3.9 分左右,试卷导向性维度的得分最低,仅有 3.77 分。究其原因,可能与区域学业成就评价大多采用传统纸笔测查、评价形式过于单一有关。试题导向性偏低会直接影响试题在反映学科发展新思想上的信、效度。

国际大型测验项目的经验表明,采用多元化评价形式是科学考查学生能力的有效途径之一。

除了传统纸笔测验,还有人机交互计算机测验、口语交际测验、现场操作测验等实践型测验,而基于该测验形式的表现性评定是当前国内外研究的热点问题。它在测查学生的高级思维能力和综合运用所学知识解决实际问题的能力、激发学生的学习动机以及优化教学过程方面有显著作用,同时也为当前的考试和评价改革提供了新的思路。^[15]

从试题与课程标准、细目蓝图的匹配程度评定结果来看,各维度平均分为 4.5 分左右,整体处于较高水平,但应该指出的是,还有 31.3% 的试卷没有细目蓝图。由表 1 可知,课程标准匹配程度较之细目蓝图而言,后者的得分更高,究其原因

可能源于细目蓝图的依据比较具体;而内容目标匹配程度较之能力目标而言,前者的得分更高,这可能是因为内容目标更具有可操作性,而能力目标比较含糊,界定不清晰。此外,结合专家具体建议进行文本分析可知,部分试题还存在高于课程标准要求或超出课程标准规定范围,对青少年的认知思维发展规律有所忽视、死记硬背的题目偏多、缺乏对问题解决等高级思维能力方面的考查等问题。

总体而言,区域基础教育学业评价的试卷命题指导思想依据、试卷结构、试题评价、试卷实效性、试卷学科特色方面表现较好,但试卷导向性偏低。试题整体与课程标准的匹配程度处于较高水平。

(二)分学科结果

由表 2、表 3 可知,各学科试卷评定和试题评定得分由高至低依次为数学、语文、英语、科

表 2 各学科试卷评定结果表

学科	总分	试卷命题指导思想依据	试卷结构	试题评价	试卷导向性	试卷实效性	试卷学科特色
语文	4.05	3.95	3.97	4.27	3.84	4.14	4.13
数学	4.40	4.39	4.24	4.42	4.29	4.55	4.42
英语	3.76	3.74	3.70	3.92	3.51	3.94	3.66
科学	3.64	3.57	3.64	3.89	3.32	3.70	3.43
品德与社会	2.85	2.81	2.75	3.40	2.65	2.55	2.65

区域基础教育学业评价与课程标准一致性的本土化研究

表3 各学科试题评定结果表

学科	与课程标准		与细目蓝图	
	内容目标的	能力目标的	内容目标的	能力目标的
	匹配程度	匹配程度	匹配程度	匹配程度
语文	4.76	4.65	4.74	4.66
数学	4.97	4.95	4.91	4.76
英语	4.56	4.37	4.73	4.72
科学	4.22	3.50	4.76	4.35
品德与社会	3.26	3.06	3.10	3.22

学、品德与社会。其中,数学学科在各维度指标上的得分最高,在总评分上达到4.4分,在与课程标准和细目蓝图匹配程度上的得分也均在4.7分以上,而品德与社会学科的得分最低,总评分仅为2.85分,与课程标准匹配程度的平均得分也仅为3分左右。造成这一问题的原因可能较为复杂,学科课程标准的清晰可操作程度、学科教师的师资力量、课时量的保证、学科测查的熟练程度等都可能影响学科评价工具的科学性和一致性。

结合各学科专家建议进行文本分析可知,区域各学科试题命题者均具有一定的命题经验和技巧,体现了命题的专业性和科学性。例如,某区小学数学试卷中有三道题考查“图形与几何”领域的知识点,其中,前两道题属于基础知识理解和掌握的范畴,第三道题则要求学生基于给定的条件自己设计方案,即要求学生对基础知识进行灵活运用,创设情境解决实际问题,这对于学生

问题解决等高级思维能力有所测查,体现了一定的开放性和灵活性。

此外,部分学科借鉴国际大型测验项目PISA的测试理念,联系学生的实际生活经验,以非连续性文本^①阅读的形式测评学生的阅读素养。例如,某区小学语文试卷中有一道题是阅读一幅有关食品安全的宣传画,要求学生基于这个非连续性文本所呈现的内容,提取信息,做出解释和评价。学会从非连续性文本中获取所需信息,并得出有意义的结论,是现代公民应具备的阅读素养,而此题恰恰对这一方面进行了考查,其测试理念与近年来国际的测试理念相符,有助于提高学生的实际生活能力。

总体而言,区域基础教育学业评价试卷评定结果在语文、数学、英语学科的表现较好,而在科学、品德与社会学科的表现有待进一步提高。同时,区域各学科在具体试题命制方面有一定的专业性和科学性,体现了国际测验的理念和趋势。

(三)分区域试卷评定结果

由图2的试卷评定结果可知,总体而言,各区域试卷结构维度的得分普遍偏低,试题评价维度的得分普遍较高;具体而言,01、02区域的表现较好,而15、16、17区域的表现较差,区域之间试卷评定结果的分布较不均衡。

由图3的试题评定结果可知,总体而言,各

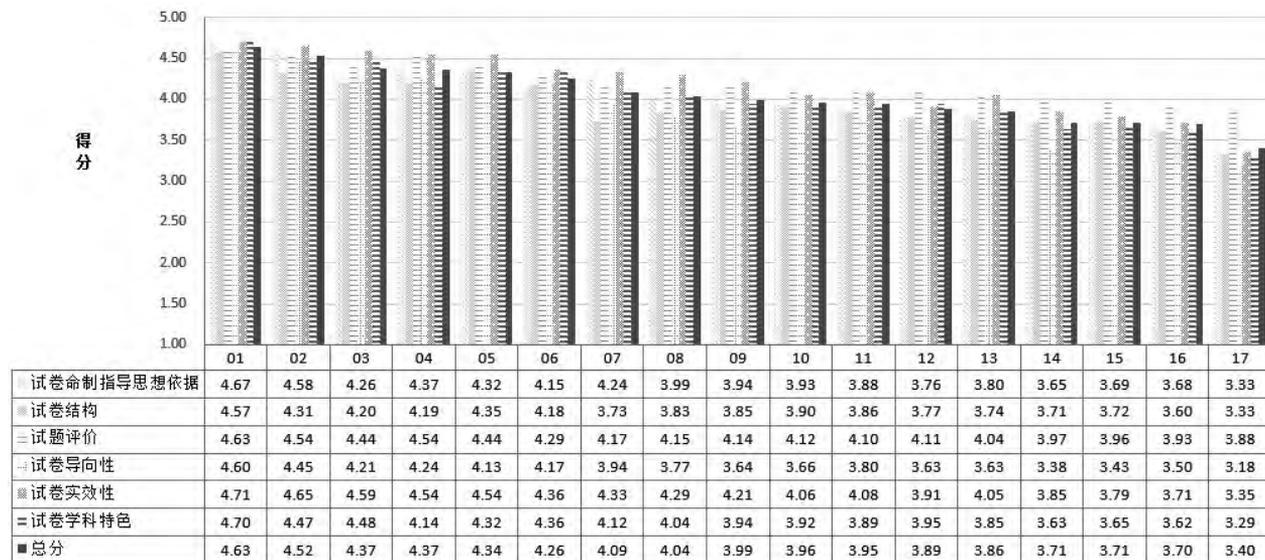


图2 各区域试卷评定结果图

区域基础教育学业评价与课程标准一致性的本土化研究

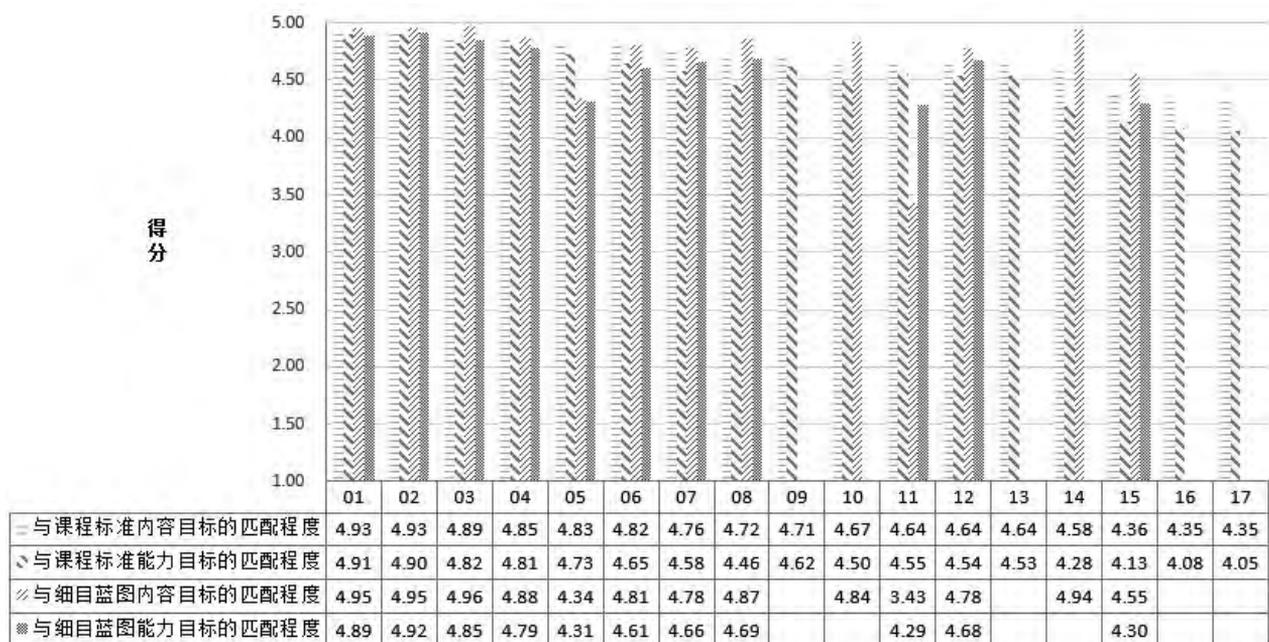


图3 各区域试题评定结果图

区域试题与内容目标的匹配程度较高,与能力目标的匹配程度较低。具体而言,01、02区域的表现较好,而16、17区域的表现较差,且09、13、16、17区域的所有学业评价工具均没有细目蓝图,区域之间试题评定结果的分布较不均衡。

此外,本研究发现,北京部分区域没有相应的评价框架和细目蓝图,或者细目蓝图编制较为简易且不科学,无法体现试卷命制的依据和思想,这一方面使得测试工具与细目蓝图的匹配程度无法考量,另一方面也使试卷内容评定在试卷命制依据和思想维度上的得分偏低。反之,我们也可以看到,对于有细目蓝图的区域而言,其学科试题与细目蓝图的匹配程度较高,这表明各区域均能很好地基于可操作性的目标进行评价工具的研制。可见,将课程标准的内容目标和能力目标具体化,编制更细致和科学的细目蓝图作为命题依据,是开展学业评价的核心基础。

三、思考与建议

(一)贯彻落实国家课程标准,研发学科学业标准

研究表明,北京各区域大部分学科基础教育学业评价与课程标准的一致性程度较高,部分学

科则有待加强。国家课程标准是义务教育教学质量评价的基础和核心,贯彻执行国家课程标准要求,是保证义务教育均衡公平发展的根本。因此,各区域教研部门和学校教师应深入理解课程标准,明确其具体内涵,注重课程标准与实际生活的联系,并将其渗透到评价工具的编制中,同时,要尽可能地采用市级专家建构的评价框架,做到资源的充分利用。

针对当前各区域基于标准命题的不明确性,以及框架蓝图等环节缺失的问题,市级层面有必要建立统一、清晰的课程标准,结合实际情况,开发各年级学业标准,开展市级培训解读,使各区域各学科教研部门、一线教师能更加明晰课程标准,为更好地指导教师教学、学生学习服务。

(二)规范命题流程环节,保证命题科学有效

基于区域基础教育学业评价与课程标准一致性的调研结果可知,各区域在命题环节上存在很大差异,有些区域命题环节比较完备,而有些区域仍然停留在凑题评测阶段。因此,各区域在以后的评价工具研制中,要保证命题环节的完整性,提升各区域命题流程的科学性、规范性,例如,成立命题小组并明确其职责,增强试卷的规范性,杜绝试卷的科学性错误,进而使试题的描

述更加规范、准确,等等。

对于命题制度和流程不规范的区域,可推行审校制度,一方面,要有评价工具的预测环节,编制细目蓝图,规范图表及语言表述;另一方面,要重基础,宽覆盖,难易适中,杜绝低级错误。此外,各区域可从对命题质量的追踪评价、审查、专家评定、一致性分析、效度研究等方面,提升和保证命题把关的有效性,力求使试题命制在信、效度层面确保其科学性;在应用层面要符合课程标准的要求,遵循学生思维的发展规律;在功能层面要使评价工作更加科学、有效,对教师“教”和学生“学”的反馈指导更具针对性。

(三)增强命题专业化培训,提高教师的命题能力

命题技术是整个评价工具开发编制的核心技术环节。虽然有研究表明区域基础教育在试题评价方面表现较好,体现了一定的专业性、科学性和国际视野,但具体到各个区域还存在很大差异,部分区域仍存在学科命题技术方面的问题,需要进一步提高命题的专业化程度。

因此,各区域教研部门和学校教师有必要把握命题原则,研究命题技术;市级教研部门应做好教师的命题技术培训工作,提高其命题能力。例如,细目蓝图的编制需要依据考试大纲,结合试卷命制的依据和思想,将内容标准、知识点明确地渗透其中,进而为试题命制打下基础,“蓝图”中应包括知识与能力考查的双向细目及分值分配比例,如主观题、客观题、评分标准的命制规则和方法等。上述任务的完成均需命题者具备一定的专业性和技术性,只有对教师进行专业化的培训,才能使其更好地为命题服务。

(四)采用多元化评价方式,关注高级思维能力的考查

区域基础教育学业评价与课程标准的一致性研究表明,试卷导向性偏低,评价方式多集中于纸笔测试,此种单一化评价方式在一定程度上易造成对学生高级思维能力考查的缺乏。同时,当前各区域基于国家课程标准的评价,主要还停留在基本的认知领域层面,更多关注的是学生的

知识点记忆、简单应用和实践层面,对于高级思维能力的考查仍有待进一步加强。因此,各区域有必要开展多元化评价方式,提升对学生高级思维能力考查的关注度。

一方面,各区域可根据实际评价内容,有针对性地整合多种评价方式,特别是对于实践类项目的考查,例如语文口语、英语听说、科学类实验操作等,应综合多元化测查手段,包括使用面对面交流、人机对话、实际操作和现场演示的方式,采用表现性评定方法,全面科学地了解进而掌握学生的能力发展状况。另一方面,在常规命题环节的基础上,各区域可整合多方力量和资源,合力推进对学生高级思维能力的测查。以PISA于2012年开始测评学生在具体情境中的问题解决能力为例,这种高于学科之上的综合能力的测评,需要市级教研部门和区域教研部门、一线教师共同研究探讨,更多关注学生不同思维水平、不同认知加工方式的评价与教学。

注释:

①与以句子和段落组成的“连续性文本”的阅读材料相对,多以统计图表、图画等形式呈现。它的特点是直观、简明、概括性强、易于比较,在现代社会中被广泛运用,且与人们的日常生活和工作须臾不离,其实用性特征和功能十分明显。

参考文献:

[1] Herman J. L., Webb N. M. & Zuniga S. A. Measurement Issues in the Alignment of Standards and Assessments: A Case Study [J]. Applied Measurement in Education, 2007, 20(1).

[2] Achieve, Inc. An Alignment Analysis of Washington States College Readiness Mathematics Standards with Various Local Placement Tests [M]. Washington, DC: Achieve, Inc. 2006.

[3] Resnick L., Rothman R., Slattery J. & Vranek J. Benchmarking and Alignment of Standards and Testing [J]. Educational Assessment, 2004, 9(1).

[4] Porter A. C.. Measuring the Content of Instruction: Uses in Research and Practice [J]. Educational Researcher, 2002, 31(7).

[5] 邵朝友,张斌,王少非.论学生学业成就评价与课程标准的一致性[J].教育研究与实验,2011(6).

(下转第39页)

美国明尼苏达州教师评价新模式及其启示

观地评价学生对学习目标掌握到何种程度,而且还注重学生在学习过程中的参与程度,从而促使教师不断修改教学实践。该评价不仅关注学生的认知领域(如学生的自我调节、明确学习目标等),还关注学生的学习情感领域(如是否乐于参与活动、是否愿意参与小组讨论等)和动作技能领域(如行为表现、同伴互动等)。利用学生的学习成果对教师进行评价无可厚非,但在我国,“学生的学习成果在许多学校和地区表现为运用纯粹的学生期中、期末或升学的考试分数来评价教师”^[7]。这就会导致教师过分追求分数,抹杀了学生的天性,忽视了学生的学习兴趣和态度等动机等非认知因素。因此,我们应从多种渠道收集有关学生学习成果的信息,如学生的身体健康水平、学生的人际交往能力、学生的道德修养、学生的业余爱好和学生成绩等对教师进行评价。一个优秀教师培养出来的学生应该是全面发展的,只有全面收集学生信息并加以科学利用,才能对教师做出客观而公正的评价。

参考文献:

[1] Minnesota Department of Education. The Teacher De-

velopment, Evaluation, and Peer Support Model Implementation Handbook [EB/OL]. http://education.state.mn.us/mdeprod/idcplg?IdcService=GET_FILE&dDocName=049729&RevisionSelectionMethod=latestReleased&Rendition=primary.pdf, 2016-03-22.

[2] Minnesota Department of Education. Performance Standards for Teacher Practice Rubric [EB/OL]. http://education.state.mn.us/mdeprod/idcplg?IdcService=GET_FILE&dDocName=049727&RevisionSelectionMethod=latestReleased&Rendition=primary.pdf, 2016-03-05.

[3] Minnesota Department of Education. Student Learning Goals Handbook [EB/OL]. http://education.state.mn.us/mdeprod/idcplg?IdcService=GET_FILE&dDocName=022053&RevisionSelectionMethod=latestReleased&Rendition=primary.pdf, 2016-03-22.

[4] Minnesota Department of Education. Content Offerings—Principal and Teacher Development and Evaluation [EB/OL]. http://education.state.mn.us/mdeprod/idcplg?IdcService=GET_FILE&dDocName=MDE034299&RevisionSelectionMethod=latestReleased&Rendition=primary.pdf, 2016-04-05.

[5] 米锦平, 代建军. 当前我国中小学教师绩效评价的问题及反思[J]. 教育科学研究, 2011(8).

[6] 高鹏, 美、加等国表现性评价制度对我国中小学教师评价的启示[J]. 上海教育评估研究, 2012(3).

[7] 王景英, 梁红梅. 当前美国中小学教师评价的特点及其启示[J]. 外国教育研究, 2002(9).

责任编辑/林洁

(上接第32页)

[6] 崔允漷, 王少非, 夏雪梅. 基于标准的学生学业成就评价[M]. 上海: 华东师范大学出版社, 2008.

[7] 刘学智, 马云鹏. 美国“SEC”一致性分析范式的诠释与启示——基础教育中评价与课程标准一致性的视角[J]. 比较教育研究, 2007, 28(5).

[8] 范立双, 刘学智. 美国“成功分析模式”的诠释与启示——学业评价与课程标准一致性的视角[J]. 比较教育研究, 2010(8).

[9] 杨玉琴, 张新宇, 占小红. 美国 Achieve“测验—标准”一致性分析工具的研究及启示[J]. 教育科学, 2011(4).

[10] 岳喜腾, 张雨强. 基于课程标准的学业成就评价: 韦伯模式之研究[J]. 全球教育展望, 2011(10).

[11] 曹小旭, 张庆霞. 基于标准的小学语文学业水平考试试卷质量分析——学业评价与课程标准一致性的视角[J]. 教育测量与评价, 2011(7).

[12] 刘学智, 曹小旭. 小学数学学业水平测试与课程标准一致性水平探究[J]. 当代教育科学, 2011(14).

[13] 王焕霞. 高中物理学业水平考试与课程标准的一致性研究[J]. 课程·教材·教法, 2015(8).

[14] 张迎春, 路小娜, 董扬扬. 教材栏目与课程标准的一致性研究——以北师大版《生物学》教材“学习目标”栏目为例[J]. 教育理论与实践, 2015(11).

[15] 赵德成. 表现性评价: 历史、实践及未来[J]. 课程·教材·教法, 2013(2).

责任编辑/林洁