

多元评价体系中制定评价标准的思考^{*}

谢利民 褚慧玲

【摘要】 新课程改革明确提出要构建多元化的评价体系,制定考试评价标准是多元评价体系构建的重要组成部分与完善,同时对有效实施考试,对有效评价学生学业成果,对有效促进学生健康发展等方面都有着积极的重要意义。课程标准是制定考试评价标准的依据和原则,借鉴国外考试评价标准制定与实施的有益经验,对制定考试评价标准的依据、基本原则及其特点等问题进行了阐述。

【关键词】 评价 标准 目标 表现水平

【作者简介】 谢利民/上海师范大学学科教育研究所所长、教授 (上海 200234)

褚慧玲/上海市教育考试院特级教师 (上海 200234)

我国在实施新课程改革的过程中,逐步建立起了与之相适应的课程评价体系,“建立促进学生全面发展的评价体系。评价不仅要关注学生的学业成绩,而且要发现和发展学生多方面的潜能,了解学生发展中的需求,帮助学生认识自我,建立自信。发挥评价的教育功能,促进学生在原有水平上的发展。”^[1]新的评价体系通过评价指标的多元化、评价方式的多样化,综合运用观察、交流、测验、实际操作、作品展示、自评与互评等多种方式,为学生建立综合、动态的成长记录手册,全面反映学生的成长历程。多元的评价体系更加注重量化评价与质性评价相结合、终结性评价与形成性评价相结合、自评与他评相结合、内部评价和外部评价相结合。

构建多元化的评价体系是课程改革过程中的一个关键。在改革的实践中我们也感觉到,尽管随着课程改革的不断深化,教学评价也在向着多元化方向发展,强化了关注过程的表现性评价等,对学生的学业评价已在不仅仅限于纸笔考试的评价方式;然而,纸笔考试毕竟还是评价学生学业水平的重要方式和组成部分。当我们在不断更新教学评价理念,实施教学评价方式与方法多元化的

同时,应该根据新的课程标准和教学目标要求对传统的纸笔考试测验方式逐步不断的进行完善,这也是完善新的课程与教学评价体系所必需的。因为许多时候和许多情况下,继承、巩固优良传统,与开拓、创新是同等重要的,有时甚至可能是更重要的。

一、制定评价标准的意义

测量是按照明确的构想和规则对人、事或物的属性或特征赋值的过程。教育测量也需要按照具体的规则对其确定的属性进行量化,教育测量的主体对象是学生的属性或特征。教学过程 中的测验,是指用于系统观察教育或心理特质或属性的一个或一系列任务,它有助于确定计划或方法是否有效,确定其是否值得继续使用。评价是根据各种观察或评价者自身背景及所受训练,而做出有价值的判断或决策的过程。^[2]

课程标准是教材编写与审查、课程实施与管理、课程评价与考试命题的依据。课程目标是根据国家的教育方针、相应年龄段教育的性质和任务及学生的身心发展规律所确定的,学生通过完

^{*} 本文为上海市教委重点研究项目“新课程背景下课堂教学本土行动研究”(06ZS25)阶段研究成果。

成规定的课程教学内容和任务所应达到的学习结果,是总的人才培养目标在某个阶段的某个方面的具体体现。

现行的国家课程标准中,没有明确的对学生通过完成规定的课程、教学内容和任务所应获得的学习成果进行描述,即没有对学生的行为进行具体描述,同时也没有制定出与行为目标相应的表现水平标准,因而不能直接将课程目标用来作为考试(特别是纸笔考试)评价的测量目标的。我国现行的课程标准的目标若直接用于学生学业评价或考试,其测量目标尚不够明确,还存在着一些不可测的因素。因此,制定考试评价标准,是对课程标准的进一步补充与完善,同时会进一步促进课程评价与教学评价的完善。

考试评价的标准应能体现课程目标,能有效

地促进教学目标的实现。考试评价所测量的学生学习成果,应能比较忠实的反映课程目标,应能表现出期望得到的心理特征体现,能描绘学生的知识、技能等在某些状态下的可以测量评价的心理特征。因此考试评价的目标必须既明确又具有可测性。考试评价的目标一般来源于课程标准。

根据课程标准,制定能够体现课程目标的快速评价标准,它比课程目标更加具体、可测量、可评估,它应能体现课程标准中的那些行为化的基本特点。当然,为了获得考试的测量目标,必须对相关学科的课程标准进行研究,分析课程标准的总目标和阶段目标中,哪些内容是在特定的考试环境中可测得,哪些内容是在该环境中不可测的,哪些内容可以作为具体的行为目标,这些行为目标能否与确定的测量目标相联系。

表一

第一阶段学习成果	第二阶段的物理的学习成果
1. 在物理实验中能选择仪器、记录观察; 2. 设计调查研究来检验物理假设; 3. 能够通过各种信息源获取物理信息; 4. 能展示对物理概念的认识和理解; 5. 从物理数据中分析、得出结论; 6. 推广解决物理问题的方法; 7. 运用物理知识,做出涉及个人、社会和环境的决定; 8. 使用物理术语和规律交流观点和推理	1. 承担实际活动; 2. 设计调查研究; 3. 通过各种途径获取物理信息; 4. 分析和评析物理信息和过程; 5. 展示对物理知识的理解; 6. 解决各种物理问题; 7. 将物理知识与所选择的现象相联系并应用; 8. 用文字、图像和口头形式有效的交流物理观点; 9. 正确的使用物理术语和符号。 技能包括实验技能、信息技能、调查研究设计技能、交流技能。

二、国外考试评价标准例析

从当前课程改革发展看,涉及学科教育的标准大致可分为两类:一类是学科教育的内容标准,另一类是学科教育的评价标准。这两类标准分别从不同的角度规定了学科教育内容,对实际的课程设计和课堂教学都会产生重要影响。

(一)澳大利亚南澳州中学课程的考试评价标准

澳大利亚南澳州的中学课程指南中对于“学习成果”有着比较具体和明确的描述,即学生经过课程学习后,能够达到的具体行为目标。对学生的学业评价是围绕即定的学习成果进行的。澳大利亚南澳州的学业评价主要由学习成果、评价内容、行为表现等主要项目组成。^[3]其中的评价内容就与课程的内容相同。如科学领域的物理学科课程 12 年级的学习成果要求如表一。

从表一中可看出物理课程在两个阶段的学习成果是不同的,要求也不一样。第二阶段成果比第一阶段要求要高。这些学习成果的制定影响着课程的评价体系,影响对学生学习的学业评价,对学生的学业评价是围绕着该课程的学习成果进行的。

上述学习成果主要描述了学生经过课程学习之后应该能展示出的行为类型,这些具体的行为是可测得,可以评估的。事实上,对学习成果的描述无须反映学生经历的学习过程,也无须反映教师的教学过程以及学习的具体内容,而只要明确学生经过课程学习后所获取的具体的成果。

课程学习中的学业评价。根据课程的阶段学习成果,相应的学习学业评价由不同的评价元素组成,每个元素必须占一定的权重,每个评价元素又由具体的表现性评价内容项目组成。例如:物理第一阶段的学习成果由“实践任务”、“测验”和“作业”三个评价元素来评价,每个评价元素的权

表二

评价元素	评价内容	测量的学习成果	权重(%)
实践任务	确定自变量和因变量,确定常量,安全进行实验操作,选择仪器,记录数据,在表格和图中表示数据,设计实验、分析数据、得出结论	主要评价学习成果 1、2、4、5、8	30
测验	测验中必须包括回忆信息,使用物理术语、规律和符号,论证理解、应用知识,画图,分析数据和推导结论,为检验物理假设设计一个调查研究活动。	主要评价学习成果 4、5、6、7、8	50
作业	课外作业可以采用各种形式,包括论文、报告、辩论、口试、录象或电子陈述、实报告,模型或有关历史的调查研究	主要评价学习成果 3、4、7、8	20

说明:(表格中“测量的学习成果”中的 1.2.3……是指[表一]中的“第一阶段学习成果”有关内容。)

重如表二。

在上面的评价内容一栏中,主要描述了不同学习成果所反映的行为特征。这些行为特征主要集中在那些可以测量的行为。

学业评价中的行为表现。在课程评价体系中针对某一评价元素还有明确的表现性评价项目。表现性评价的标准陈述了学生将遵循的行为品质和特征。例如,针对表二中的评价元素“实践任务”,其表现性项目所对应的具体表现性行为体现如表三。

表三

表现性评价项目	表现性行为
实验设计技能	学生能否清楚的确定实验目的 学生能否清楚的确定自变量和因变量 学生能否清楚的确定作为常量的因素 学生能否清楚的设计和描述验证一个假设所需要的过程
实践技能	学生能否在一定程度遵循安全和仪器精确度的说明 学生对疑难问题的认识和反应 学生能否正确的选择仪器 学生在实验中能否进行仔细和诚实的观察 能否有效的与其他同学进行合作
表达	学生选择数据的处理和表达是否恰当 学生能否正确的使用物理术语和单位 学生表达信息是否具有逻辑性
预测和评析	学生能否有效地分析、推断、评价一个实验,并提出修改建议 学生能否确定信息源的正误 学生能否清晰的描述结果,推出有效的结论

12 年级物理课程的学业评价由考试(50%)、作业(25%)、实践任务(15%)、信息获取与表达(10%)四方面组成,括号内为该元素在总评价中所占权重。对于每个评价元素必须说明相应的评价内容、所要测量的学习成果、表现性行为的标准等。其中考试是学校以外实施的统一考试,作业、

实践任务和信息获取与表达是由学校内部的教师进行评价。

12 年级物理课程的外部考试规定用三个小时的考试对学生的学习成果进行评价。考试内容与分数的比例分别为:物理知识的获取占 15%;理解和解决物理问题占 50%;运用物理知识占 15%;交流物理知识占 20%。试卷结构有 A、B 二卷组成,A 卷:覆盖物理课程指南中规定的课程内容的 16 个主题和一些实验技能,所有问题涉及的内容都是必修的,考试时间规定为 135 分钟,权重 75%;B 卷:二个拓展问题和一个实践技能问题,考试时间为 45 分,权重 25%。考试中学生的表现性行为主要包括:知识和理解、应用、分析、评估和交流。其中“知识和理解”具体描述:精确认识和使用物理术语和概念的程度,解释物理术语和概念的有效性,推断相关物理信息的有效性。

(二)美国国家教育进展评估

美国国家教育进展评估(简称 NAEP)是全国性的评价系统,目的是通过测量学生的学业水平反映美国教育成就的进展情况。NAEP 的首次评价于 1969 年进行,其评价对象是来自全美的 4、8、12 三个年级的学生样本。NAEP“科学”部分的评价体系在保持历史连贯性的同时,也逐渐做一些调整以适应科学教育理论和实践的变化。如 1996 年的科学教育的评价框架由 4 个部分组成:学科领域(地球、物质和生命科学),知与行(概念理解、科学研究、实际推理),科学本质(科学和技术发展史、体现科学和技术特征的思维习惯、探究方法和问题解决),统一的概念(模型、系统、变化方式)。

最近推出的科学教育评价体系由科学知识、科学实践两个维度构成。科学知识的维度由物理科学内容陈述、生命科学内容陈述、地球与空间科

学内容陈述组成。物理科学包括:物质、能量、运动三个主题;生命科学包括:生命系统的结构和功能和生命系统的变化二个主题;地球与空间科学包括在空间与时间中的地球、地球的结构和地球系统三个主题。^[4]

新的科学教育评价标准中的科学内容,是以内容领域中要点和年级水平组织的方式呈现的,反映了学科在不同年龄段的具体要求,在横向和纵向两个方面呈现出学科的体系。例如,物理内容中的能量主题“能量转化和守恒”在不同年级的内容要求:

4 年级

流过通过电路的电产生磁效应;电路包含电池、灯泡或小灯泡、电池的能量转移到灯泡,再转化成使它们发光、发声、发热的能量。

8 年级

当能量从一个物体转移到另一个物体时,转移后的总能量等于转移前的总能量。例如,一个物体下落,它的势能减少了,它的速度和动能增加了。当一个物体下落过程中,它的动能转移到介质中,使介质运动并发热。核反应发生在太阳中。在行星中,发生光和作用时,具有化学势能。

12 年级

热增加原子组成元素和分子或离子组成化合物时的平动、转动、振动的能量。当原子、分子或离子的平动能增加时,物质的温度升高。加热一个晶体,其原子、分子或离子的振动能增加。当振动能量足够大时,晶体结构破坏被熔化。当物体从离地面较远处靠近地面时,物体位置发生变化,物体势能改变。化学反应要么释放能量到环境中,要么吸收环境中的能量。核反应——聚变和裂变——将很小物质的量转化成巨大能量。在封闭系统中总能量守恒。

科学实践的维度由识别科学原理、使用科学原理、使用科学探究、使用科技设计组成,并有相应的行为目标。如识别科学原理包括:描述、测量、分类观察、陈述或再认正确的科学原理、证明相关科学原理之间的联系、证明原理不同表达形式和数据模型之间的关系。

NAEP 评价体系中,除了科学知识、科学实践两个维度外,还将学生行为表现水平划分为高级、中级和达标三个水平,并用语言对不同水平进行

了描述,并且针对不同目标、不同水平层次设置了评价项目的样例。

另外,加拿大某州的科学课程评价体系主要由评价目标、评价内容、评价中的样题举例等组成;英国的某门课程的评价体系主要由详述的评价目标、评价主题、详述内容课本与其他资源等组成。虽然各有特点,但主要组成基本相同。学业评价体系主要由评价目标、评价内容、评价中样题举例、评价资源等组成。

无论怎样的考试评价,都涉及到任务的真实性、测量技术的完善性、训练有素且公正的评价、以及建立合理且有效的考试评价标准。当然,在考试评价过程中必须平衡地运用多种方法,从而可以给学生提供一个丰富、合理且可用的学业成就特征图。考试评价关注新观点、问题、文本及知识,同时其设计又引人注目并能激发学生在智力上探索新知识的兴趣。评价体系也要考虑资源、时间的重新调配,从而可设计、调试、运用、评估及有效地反馈学生的学业状况。

三、制定考试评价标准的依据

课程标准是国家教育质量在特定教育阶段应达到的具体指标,在一定的意义上说它具有法定的性质。课程标准对教材、教学和评价具有重要指导意义,是教材、教学和评价的出发点与归宿。课程标准是国家管理和评价课程的基础,也是教材编写、教学、评估和考试命题的依据,是制定各类考试评价标准的重要依据。

(一)评价标准必须努力体现课程标准所倡导的价值观。

我国正在实施的新课程改革,从知识与技能、过程与方法、情感态度价值观三个维度对课程标准进行了阐述。以往的教学大纲比较强调的是知识和技能目标,而现在实施的新课程标准在关注知识和技能的同时,更加强调了学生学习的过程、方法、情感、态度及价值观等问题。因此,在制定考试评价标准过程中,必须摆脱以往的考试过分看重考核学生知识和技能的传统观念的束缚,仔细研读和领会课程标准的基本精神与内涵,在改革实践过程中积极探索对学生学习的过程、方法、情感态度价值观评价的有效途径和方法。

(二) 评价标准必须符合课程标准的基本要求

现行课程标准提出的教学过程所要实现的课程目标,与过去的教学大纲有着很大的不同。过去教学大纲规定了教材、教学和评价的最高要求,教材、教学、评价都不能突破这一上限,若突破了则被视为“超纲”。而现在的课程标准所提出的目标要求应该是最基本的要求,是绝大多数学生经过努力都能达到的要求。因此,制定评价标准,应遵守课程标准的各项规定与要求,特别是教学所要实现的课程目标和各科所要学习的课程内容要求。

(三) 评价标准的测量目标应具有可测性

课程标准用尽可能清晰的行为动词,从知识与技能、过程与方法、情感态度价值观三方面对学生的学学习结果进行描述,但这种描述相对还是比较原则,每个目标所包含的内容与范围较宽泛,如果要对这些目标进行测量,我们仍需将课程目标分解与细化为更为具体的考试目标。考试评价的目标要既能体现和反映课程标准,又要给考试评价的使用者(命题者、教师、学生)提供清晰的界定。因此,考试评价目标既要做到与课程目标基本保持一致的同时,又要进一步将课程目标分解与细化,提高可操作性、可测性。

四、制定考试评价标准的基本原则

根据课程标准制定考试评价标准,应注意范畴的一致性、知识深度的连贯性、知识覆盖的相近和代表性的平衡。

范畴的一致性,要求考试评价与课程标准的范畴应保持一致。建立考试评价标准也是我们这次课程改革的重要配套性改革,考试评价必须体现新课程的理念和学科课程标准的要求。要充分体现课程改革所倡导的新理念,符合基础教育改革中评价改革的方向。例如,评价目标来源于课程目标,评价内容要遵照课程的内容等。

知识深度的连贯性,要求对知识内容的复杂性水平和种类进行相关描述。描述内容复杂性一般包括四个水平,最低水平需要描述信息、事实、技能的记忆和再认;其次一个水平描述概念和概念的应用,方法的理解,和概念的理解;第三个水平需要更多的推理和分析,要求学生解决问题,推

断数据、论证,建构模型,转化不同表达方式,评判证据等;第四个水平要求从各种资源中综合运用知识,以不同方式表示知识的能力,如搜集信息、分析决策,准备报告等等。

例如,加拿大安大略省的《科学》课程评价标准对表现水平的描述分为四个等级。^[5]

水平 1

只理解几个基本概念、存在严重错误、对概念只能解释一点儿;对所要求的技能和方法使用得很少、仅知道一点儿安全操作程序、靠帮助才能正确使用工具、仪器和材料;难以清楚准确地交流、很少能使用恰当的科技术语和测量单位;对熟悉事务中的科技的关系理解得很少、对科技与社会相结合的理解很少。

水平 2

能够理解一些概念、存在一些误解、能够部分地解释概念;对所要求的技能和方法能够使用一些、知道一些安全操作程序、靠一点儿帮助能正确使用工具、仪器和材料;一定程度上能进行清楚准确地交流、有时能使用恰当的科技术语和测量单位;对熟悉事务中的科技的关系有一些理解得、对科技与社会相结合的有一些理解

水平 3

对所要求的技能和方法能够使用大部分、知道一般的安全操作程序、能正确使用工具、仪器和材料,偶尔会需要帮助;一般能清楚准确地交流、经常能使用恰当的科技术语和测量单位;能理解熟悉事务中的科技的关系、能够理解科技与社会相结合

水平 4

能够理解全部的基本概念、完全没有误解、总能完成地解释概念;对所要求的技能和方法能够使用全部、完全知道安全操作程序、无需或很少需要帮助就能正确使用工具、仪器和材料;完全地进行清楚准确地交流、完全能使用恰当的科技术语和测量单位;对周围事务中的科技的关系都能理解得、能够科技与社会相结合及其产生的影响。

大多数基于标准的评价描述学生达到的水平,通常用“高级、中级、达标”来界定,或用“水平一、水平二、水平三、水平四”来界定。

知识覆盖的相近,是要求考试评价的内容覆盖面应与被测量领域的范围相近,知识范围确保

与课程标准期望的知识基本一致,包含知识面得覆盖和知识深度,可以通过测量考试评价中任务数目来实现足够的知识范围。

代表性的平衡,要求不同目标和各内容领域中的重点,应与课程标准一致。因为各个目标在不同的内容领域中所占比例不同,因此,评价中选取某内容领域中的测量目标的重点,应与课程标准中的重点相符。

另外,制定考试评价标准还需要注意,知识结构:需要根据学习理论界定内容领域;公正、包容、公平:对目标群体进行清晰描述,专家进行内容分析;内容完整性:由于评价项目数量有限,应优先确保最重要的内容被考查。等等。

五、考试评价标准的特点

我国基础教育课程改革,制定了每个学科的课程标准,其中对学科的内容领域有很详细的要求,可作为评价标准的内容领域。由此,考试评价标准的基本框架,可由考试评价的测量目标、样题、行为特征分析等部分构成。

一般来说,测量目标包括三个层次:测量目标、行为目标和表现水平标准。测量目标是第一层次的目标,它相当于测量的心理结构,是不可直接测量的,理论上应该与课程的总的培养目标相对应。行为目标是测量目标的具体行为表现,是判断评价对象是否达到课程培养目标的具体观察行为。表现水平标准是用文字表述的学生不同水平等级的标准,是判断学生行为表现优劣程度的依据,可分为几个不同的水平等级。^[6]

评价标准的内容领域应遵照课程标准中所要求的相应的学科内容。样题是针对每个行为目标及相应的四个层次水平标准制定的,用来说明达到相应水平的要求,包括客观题和主观题两类。行为特征分析主要是针对不同行为目标在每个内容领域表现出来,哪些行为目标在哪些内容领域

表现得最为典型等进行分析。

考试评价标准中的测量目标,应该能够比较清晰的反映出考生经过一定阶段学习后,所获得的最终结果或达到的目标。对于目标的描述应该是简洁、明确、清楚的。

评价标准的内容领域应遵照课程标准中所要求的相应的学科内容。样题是针对每个行为目标及相应的不同层次水平标准制定的,用来说明达到相应水平的要求,包括客观题和主观题两类。行为特征分析主要是针对不同行为目标在每个内容领域表现出来,哪些行为目标在哪些内容领域表现得最为典型等进行分析。

现在,尽管国家课程标准是最基本的要求,但不同类型的考试通常会有不同的要求。具体到某一类考试,还要考虑到考试性质、教材、教学实际等多方面的因素。每一种考试评价都应根据该类考试评价的性质和功能,编制评价项目。因此,仔细对照课程标准,对考试要求进行准确定位,了解教师对课程标准的掌握情况,课堂教学的实际水平,教师和学生对考试的认识、要求和建议,组织富有教学经验的教师和专家反复研讨,也可以为实施一个考试创造很多有利条件。

参考文献:

- [1]教育部. 基础教育改革纲要(试行). 教基[2001]17号(2001年6月7日).
- [2][美]吉尔伯特·萨克斯著. 教育的心理的测量与评价原理[M]. 南京:江苏教育出版社,2002:13.
- [3]谢利民, 褚慧玲:关注学生学习过程的学业评价. 外国中小学教育,2007(5).
- [4]褚慧玲. 美国国家教育进展评估中的物理内容与评价. 物理教学,2007(12).
- [5][加]安大略省教育和培训部. 《科学》课程标准(1-8年级)[M]. 北京:科学出版社,2006:83.
- [6]雷新勇. 大规模教育考试:命题与评价[M]. 上海:华东师范大学出版社,2006:83.

(责任校对:周巧学)

