



# 工大 道

2015年第二期（总第9期）

浙江工业大学教师教学发展中心编

2015年4月20日出版

## 本期要目

### 本期聚焦·学业评价

理解课堂中的形成性评价	1
形成性评价在高校课程评价中的应用	7
美国著名大学教学评价的内容特征	13
影响评估效果的因素	18
教师自行设计的考试	24
评判分数	45

### 工大教风·教师论文

基于多元智能理论的大学生学业评价模式构建	54
----------------------	----

## 理解课堂中的形成性评价

(美) Ellen Osmundson 著 何珊云 王小平 译

**摘要:** 形成性评价是一种能促进学生学习的有效策略, 本文的核心问题正是如何理解课堂中的形成性评价。为此, 本文将首先介绍目前美国教育系统中存在的一些问题, 并就此问题提出一个可能的解决方案——形成性评价。在简要的背景回顾和文献综述之后, 文章将提出我们构建的形成性评价工作模型。接着, 本文将从我们围绕形成性评价设计的教师专业发展项目, 以及我们自身的探索出发, 讨论形成性评价面临的一些问题。最后便是根据这些新问题, 分析如何进一步理解形成性评价。

**关键词:** 形成性评价 课堂 工作模型

### 一、当前的问题

各种有关美国教育状况的研究报告中无不充斥着这样一个令人沮丧的消息: 在最差的一批学校中, 有50%的学生未能完成高中学业而退学, 大学教师悲哀地发现, 他们不得不对新入学的学生进行数学和写作课的补习。与此同时, 企业界则在抱怨, 高校毕业生远没有为进入职场作好准备, 他们认为, 这些毕业生缺乏解决问题和批判性思维的能力。中小学教师同样面临着各种挑战; 学校、学区和各州都在大规模地消减所有教育项目预算, 因此, 旨在加强学习和提升概念知识、教学知识和评价知识为目的的教师专业发展几乎从“教师的生活”中消失殆尽。显然, 我们正处于一次教学与学习的危机之中。

但是在这一教育深陷困境的黑暗时刻, 我们仍能看到一丝令人鼓舞的亮光。2001年颁布《有教无类法》(No Child Left Behind Act of 2001), 以及2009年国会以立法方式通过“力争上游”(Race to the Top)计划, 均激发了人们对于运用评价去测量和促进学生学习的兴趣。课堂层面评价的证据是令人信服的: 统计数据显示, 教师持续使用评价来指导和引领教学——也就是形成性评价——对学生学习有非常显著的提高。这种连续的评估过程缩小了学生当前的认识水平与期望目标之间的差距, 对于帮助那些表现不佳的学生取得学业进步尤其有效。然而最新的研究结果显示, 如何实施高质量的形成性评价, 使其对学生的学习发挥重要作用, 依然充满着挑战。

我的同事, 同样来自CRESST的Joan Herman博士曾在其所写的文章中提出, 当前急需深刻理解评价的三重含义, 即学习的评价(assessment of learning)、

促进学习的评价(assessment for learning)和作为学习的评价(assessment as learning)，并在此基础上形成一套连贯、全面、可持续的评估系统。我的这篇论文将更多关注形成性评价，它正是旨在支持学生学习、提高学业成就的全面评价系统的组成部分。

## 二、形成性评价的背景和已有研究

通过梳理250多项关于教学与学习的研究成果，1998年，Paul Black和Dylan William最终发现，在所有的教育性干预研究报告中，大家最关心的问题正是课堂教学中运用形成性评价策略对于学生的促进作用。报告显示，这种影响水平普遍在0.4到0.7之间，其中对于表现不佳的同学的影响特别强烈。

国家研究委员会(National Research Council, NRC)2001年推出报告《知道学生知道什么》(Knowing What Student Know)，强调教学与形成性评价之间关系的重要性，并指出：“如果教学和评价紧密相连，学生将会学到更多。在课堂中，给学生提供详细信息，尤其是详细评价他们的作业质量，或向他们提供改进方法，乃是最大限度促进学生学习的关键。”

关于形成性评价在促进学生学习方面的重要意义，各种相关研究已达成广泛一致。本研究将从动机理论、元认知、自我调节、课堂评价、教师话语实践以及反馈等不同的理论视角出发，认为在实践中，形成性评价是一个动态的、互动的过程，包括：

- 教师和学生分享并理解学习目标和成功标准；
- 通过教师评价、自我评价和同伴互评，对学习情况进行持续性监控；
- 教学反馈与及时调整，以满足学习者的需求。

然而最近的研究表明，在学习以及如何将形成性评价实践融入教学以促进学生学习方面，教师恰恰正面临着巨大挑战。

## 三、形成性评价的工作模型

图1展示了当前我们正在使用形成性评价工作模型(working model)。我们同意并感谢CRESST同事Margaret Heritage设计的“形成性评价环”(the formative assessment cycle)。在Margaret Heritage的设计中，形成性评价共分十个环节；各环节之间的关系是动态且互动的，也就是说，一个环节的变化将会导致过程中其余部分的变化。接下来将对每个环节进行简短介绍：

- 学习进程(learning progressions)。学习进程由学习经验描述构成，

旨在交代特定内容领域或某一时间段的学习将如何发展/进步。因此学习进程可以提示、引导学生接近更直接的学习目标和成功标准。

——学习目标(learning goals)。学习目标描述了学生在教学过程(不是一天的学习活动)中将学到什么。学习目标必须清晰、可控,符合学生的学习需要,并与成功标准相一致。学生不仅理解学习目标,而且彼此可以交流学习目标。

——成功标准(criteria for success)。成功标准是学生表达学习目标的基本方式。学生可以通过解释、示范、演示、问题解决等,证明他们的目标。有效的成功标准必须清晰,与学习目标一致,公平且毫无偏见。成功标准还应方便学生交流和理解,并且可以让学生通过案例来证明自己达到了标准。

——获取学习证据(elicitting evidence of learning)。通过特定策略和方法,系统搜集学生学习过程中出现的各种证据。所采取的策略与方法必须与学习目标、成功标准保持一致。此外,所搜集到的证明必须能够反映重要概念和技能的掌握情况。

——解释证据(interpreting the evidence)。指教师根据证据信息分析学生学到哪里了,进而明确哪里需要增加教学。

——诊断差距(identifying the gap)。指教师分析学生离达到学习目标和成功标准还有多少距离。

——给予反馈(feedback)。对学生的学习给予反馈是促进学生不断进步的关键。详细且具有针对性的反馈信息能够支持学生的学习,这些信息提出了怎样才是成功以及需要改进的部分,而不是仅仅列举解决问题的所有措施。

——教学调整(instructional modifications)。接下来就是调整教学来支持学生的学习。教师可以通过提供额外的学习经验,鼓励学生参加讨论,或是引导性地发问来帮助学生改进学习。

——搭建支架(scaffolding new learning)。给新的学习搭建支架包括给学生提供适当水平的支持,从而使学生能够独立地说、写、制作等。

——缩小差距(closing the gap)。缩小差距是指推动学生沿着他们的学习轨迹缩小他们目前知道的、说的、做的和学习目标之间的差距。这是教学和学习最为关键的一步。



图1 形成性评价工作模型 (working model)

课堂文化是学习发生的环境，也是我们模式中的最后一个组成部分。课堂是一个合作的地方，在课堂中，学生清楚地理解学习目标和成功标准，获得具体和特定的反馈，并有合适的时间与机会通过自我反思或同伴评价来改进自身的学习。课堂氛围的质量在支持学习方面发挥着核心作用。学生需要相信，课堂是属于他们的，教师和同伴在课堂中可以帮助他们学习。教师必须进行明确的教学，并举例说明课堂期望，这种期望反映了积极的社会行为，不具威胁的互动行为和合作性的日常工作。

劳瑞·谢帕德 (Lorrie Shepard) 在2000年AERA的大会发言上，提出有关课堂文化、形成性评价和学习的几个问题：

“课堂文化该如何转变，才能使学生最终不需要假装有能力或在测试中表现得很好，而是使学生真正地学习？我们究竟该做些什么来改变课堂，才能使学生和教师将评价作为思考和彼此帮助的开端？为了实现这种转变，我们必须使评价在学习中更为有用，更能帮助学生和教师……”

#### 四、形成性评价的新近研究

在过去的七年中，我们CRESST 团队一直在研究形成性评价，同时围绕形成性评价开发教师专业发展项目。接下来便是概述我们这些年来在教师形成性评价能力，以及理解教师形成性评价实践方面，通过不断努力所取得的发现。

##### (一) 全州范围的行动

第一个项目是一项持续多年努力的行动，包括州教育部工作者、领导团队、职前教育培训者、行政人员和教师在内的不同层次人员的参与以及广泛的合作。这些参与者出席了一系列的“面对面会议” (face-to-face meetings)，会议

的重点议题是建立形成性评价知识，理解形成性评价过程中的不同构成要素。在此之后，我们还组建了网络研讨会 (Webinar)、专业学习共同体 (Professional Learning Communities, PLCs)，并为参与者在课堂、学校或地区情境中尝试形成性评价新策略，提供了许多机会。年复一年，越来越多的参与者加入了我们的形成性评价活动，我们的形成性评价活动也与“推广国家标准”、“反馈教育干预”等其他许多州级行动融为一体了。

透过这些项目，有几点发现值得一提。如果想要在全州范围培养教师的形成性评价能力，就必须实施并且长期致力于变革。当不同的参与者一起学习形成性评价时，必须珍视持续的多方合作，尽管多方合作做起来比较难，但它的确有价值，因为它可以让大家看到不同的学习与试验水平。最后，对任何大规模的变革行动来说，计划、反馈和修改都是至关重要的。

## (二) 专业发展计划

第二个项目通过提供学术机构、丰富的内容 (科学) 知识以及有关评价知识学习的机会，致力于为 K-12 阶段的教师提供形成性评价知识领域的专业发展。并采用“形成性评价档案袋”的方式来记录教师的学习情况，以此作为教师形成性评价知识和实践发展的数据和信息资源。教师以形成性方式使用现有科学课程和评价方式，深化他们的评价知识和实践。学区的专业学习共同体为教师提供机遇与合作，塑造他们的领导力。此外，每一学区都有机会接受形成性评价专家的指导。在参与项目的教师尝试用新的策略和方法来理解学生的学习时，专家们往往会提供支持和帮助。

通过这一项目的实施，我们开始理解帮助教师构建专业知识的重要性，同时也要看到丰富的学习机会可能会带来的优势和劣势。学术机构为教师提供了多样的机会，帮助他们构建学科内容和形成性评价的知识，但对于教师来说，如何在日常课堂活动之外尝试这些复杂的想法和策略，往往是十分沉重的负担。随着教师更好地理解课堂活动如何成为学生学习信息的一个重要来源，并用新的眼光来检视材料，他们发现，当学习目标和成功标准相一致时，对包括评价在内的现有材料的使用就成为类似寻宝的行动。最后，对于参与者来说，幼儿园到 12 年级老师之间的合作都是有利的。通过与知识更加渊博的高中教师交谈，小学教师加深了他们的科学学科内容知识，同样地，通过与小学教师互动，高中教师丰富了他们的教学方法和教学内容知识。

### (三) 嵌入式评估系统

最后一个项目是小学三、四年级的科学老师使用“嵌入课程式形成性评价系统”的随机对照组研究。实验组的老师选择探究型课程，采取前测/后测、科学笔记、总结性检查表和编码指南等工具对学生的学习进行分析，同时，教师与学习小组中的其他教师开会，为学生学习提供支持。控制组教师则“照常”进行的科学学科课程的教学和评价。我们使用在概念上与高质量的形成性评价实践相一致的理论工具来收集关于教师运用形成性评价的数据。数据来源涵盖对教师教学前/后学科内容知识(content knowledge)和教学内容知识(pedagogical content knowledge)的调查，记录每周教学实际情况的博客，以及课堂观察和访谈。

我们的研究暂时进行到这一步，数据分析正在进行中。例如，我们发现通过对具体学科的前/后测的结果，实验组的教师在使用嵌入式评价系统后，提升了他们的学科内容知识和教学内容知识，这些或许会最终促进学生科学学科的学习。

从上一个研究项目中，我们体会到了实践的重要性；而在这个研究中，我们发现在使用嵌入式评价系统的第二年后，实验组教师能更清晰地理解检查学生学习所用工具和实施方案；能更有规律、更有系统地分析学生学习；能更频繁地给予学生反馈，提供给他们具体的改进意见；以及能更坚定地参加下一步的工作，那就是为学生的学习提供帮衬和辅助。此外，教师们还表示，他们对于学习目标和学科单元学习的成功标准有了更深的理解。另外，通过课堂观察、访谈和教师电子记录，这项研究表明教师知识上的细小转变可能代表着形成性评价实践中的显著改变。这一改变令人鼓舞，标志着我们视形成性评价为改进教学之良药的信念是确证有效的。

## 五、结论及后续步骤

本文从当前美国教育正在面临的挑战开始，即“我们如何才能改善美国学校的教学质量？”我认为，形成性评价是强化学生学习和改善教学质量的一条路径。提高教师运用形成性评价的能力，以此来增加教学有效性是一个极具挑战的命题。然而，如果我们认真地肩负起为所有学生提供优质教学责任，那么把立足于并根源于师生日常生活的形成性评价，作为改进教学的一种方法，仍是具有吸引力和说服力的。

(本文摘自《全球教育展望》2012年第4期)

## 形成性评价在高校课程评价中的应用

林 静

**摘要:** 随着我国素质教育的全面推广和普及,形成性评价开始逐渐应用于高校课程评价中。形成性评价在高校课程中的应用应坚持评价主体的多元化、评价方式的多样化、评价内容的发展性、评价标准的弹性设置等基本原则,在评价中应建立评价学生全面发展的指标体系,形成性评价与终结性评价相结合,加强形成性评价制度建设等。

**关键词:** 形成性评价; 高校课程; 应用策略

20世纪90年代末,随着我国素质教育的全面推广和普及,形成性评价以其更多地关注学生的学习过程、学习方法、情感态度,更能全面客观地评价学生的综合素质等优势,开始逐渐应用于高校课程评价中。本文就形成性评价在高校课程评价中的应用略陈管见。

### 一、形成性评价的理论基础

形成性评价是美国课程评价专家斯克瑞文于1967年首先提出,随后一直在西方发达国家的教育评价中被广泛推行。形成性评价“是通过诊断教育方案或计划、教育过程或活动中存在的问题,为正在进行的教育活动提供反馈信息,以提高正在进行的教育活动质量的评价”。从实施过程来看,形成性评价与中医的“望闻问切”诊断方法有异曲同工之妙,通过“评价诊断——反馈信息——修正问题——提高质量”等阶段,通过对教学各环节的实时监控,形成对教育者和受教者即时性的影响效应,切实达到提高教学质量和人才培养质量这一根本宗旨。作为教学环节主体的师生,可以通过形成性评价的诊断,及时反思,不断完善,保证教学过程的完整性,提高质量,确保实效性;与此同时,形成性评价在机制上“诊断、反馈、修正”三位一体,确保推动“评价不是为了证明,而是为了改进”。

美国教育学家布卢姆的形成性评价教学实践和研究表明,形成性评价不仅能改进教师的教学方法、提高课堂教学效果,更能挖掘学生的学习潜力、促进学生的智能发展、提高学习效果。其主要功能表现在:

第一,有效调整教学活动。教师在课前明确规定每个学习阶段的学习目标及评价项目,以单元学习开展教学,并根据评价结果及时调整教学目标和教学活动。

第二,及时强化学习效果。学生能在每个学习阶段明确自己的目标是否达到,通过教师的评价反馈明白学习过程中存在的问题,自主调整学习目标,继而调

动学习积极性，增强自信心，起到强化学习效果的作用。

第三，高效解决问题症结。根据对存在问题的分析，教师及时地调整教学目标，改进教学方法，同时给学生提供辅导和帮助，使他们自觉地调整学习计划，改进学习方法，完成学习任务，提高学业成绩，最终获得教与学的协同发展。

## 二、形成性评价在高校课程评价中的应用原则

高校课程教学本质上是以教材为载体的，在教学上，大学课堂更注重学生自主性的发挥与师生互动，因此如何调动师生共同参与教学和学习的分析、讨论是至关重要的。形成性评价通过评价主体的多元化、评价方式的多样化、评价内容的发展性和评价标准的弹性设置等特点来保证师生对教学及学习的有效反思，提高他们对形成性评价的参与积极性，从而进一步激发学生的学习动力，增强学生学习的主动性、积极性和责任感，使得学生端正学习态度，改进学习方法，提高学习能力。

### （一）评价主体的多元化

“评价主体的多元化”是指评价活动可以由多人来主持，使学生从传统的评价客体转变为评价主体，提倡多主体参与评价，鼓励学生本人、同学、实习单位指导教师等参与到评价中，建立主体多元化的评价体系和评价制度。

#### 1. 教师评价

在形成性评价中，教师的评价仍将发挥重要作用，但不再是唯一评价。每位教师都希望运用恰当的评价来促进学生的发展。教师作为评价活动的策划者、组织者和参与者，主要是完成对学生的评价。体现在考查学生的学习态度，考查学生在学习过程中表现的学习自主能力、学习自律能力、学习发现能力、学习拓展能力等多种学习能力。

#### 2. 学生评价

学生通过自评和互评，能及时发现自己或他人在学习中存在的问题，认真做好并保存好相关学习记录，培养良好的归因品质，同时在自评和互评中进行总结、反思，不断改进学习方法，提高元认知水平。

#### 3. 实习单位指导教师评价

在社会实践、专业实习、毕业设计、毕业实习等实践环节中，通过实习单位相关指导教师的评价，了解学生的理论应用实践能力、操作动手能力，可以清楚地找出学生所欠缺的知识和能力。这样，有利于高校人才培养更好地与社会需求

相一致。

## （二）评价方式的多样化

评价方式的多样化体现在两方面，一方面，评价主体不一致，其评价方式也不一致，另一方面，同一评价主体，其评价方式也可多种多样。高校的课程教学过程是丰富生动的，其评价方式也应强调丰富多样和优化组合。教师制定课程考核大纲，根据教学内容、教学活动安排和师生实际情况，灵活选择和组合。包括教师观察、座谈、访谈、检查、测试，学生自评、同学互评、小组讨论、小组测评、网络测评，实习单位指导教师观察、座谈、测评等。评价可以采用量表打分形式，也可以定性描述评价，可以小组讨论集体测评，也可以由师生单独测评。具体采用哪种评价方式，要根据教学内容和师生实际情况来进行有效选择和组合。

## （三）评价内容的发展性

从评价内容上来说，形成性评价更注重评价内容的发展性。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. 高校理论课程的评价内容

（1）相关学习内容测试。以采用教师评价为主，命题笔试形式进行，包括分阶段的单元测试、期中考试和期末考试。主要检测学生在某一阶段对专业知识的理解、运用和分析与判断能力。教师必须根据学生在检测中的真实成绩给予客观公正的评定。

（2）日常学习表现观察。可采用教师评价或学生评价。日常学习表现主要包括考勤情况、学习笔记、学习记录、课堂表现等。主要检测学生的学习态度、学习纪律和自我管理能力。对于平时无故旷课、随意迟到早退的学生扣除相应的平时分，对于在课堂上表现积极、主动提问的同学给予加分鼓励，以此督促学生更好地进行平常的学习活动，取得学习成效。

（3）平时作业检查。可采用教师评价或学生评价。主要检测学生在每一阶段的学习成效。在日常学习中，教师应当适时布置适量作业，一般要有对新学知识的检验，可以通过课后练习来完成，更要有对下节课教学内容相关的阅读和问题，提交一份小作业，并指出自己认为难以理解的部分。相对测试来说，平时作业能及时地检验教师的教学质量，因此，高校教师应认真对待平时作业这一环节，

并从学生作业中找出教学漏洞，进一步改进教学方法。

(4) 专题论文撰写。以采用教师评价为主。主要检测学生的知识面与归整能力、分析与解决问题能力、逻辑思维能力和书面表达能力。教师可根据不同的教学内容设置相应的论文题目，由学生个人或团体完成，学生通过收集资料、综合分析、提出自己的独到见解，最后以论文的方式上交。

(5) 合作学习或小组讨论总结。可采用教师评价或学生评价。主要检测合作技能、团队精神、沟通能力、信息获取和分析技能、知识学习能力和沟通协调能力。通过团队学生间的分工合作或讨论及共同完成学习任务的情况，全面评价学生知识与能力发展水平。教师还可根据不同的课程，组织各种教学活动，进行相应的评价，如交流学习心得、开展课题研究、问卷调查、演讲、辩论，等等。不管采取哪种形式，有一点是肯定的，评价的内容应更加关注学生的成长成才，更多关注学生的能力发展。

## 2. 高校实践课程的评价内容

实践课程的形成性评价是以理论课程的形成性评价为基准的，实践课程不能与理论课程完全独立分开，要在理论课程的基础上，以学生的发展为前提，进行实践课程的形成性评价。

(1) 日常实践观察。可采用教师评价、学生评价或实践单位评价。主要检测学生的理论知识掌握程度、理论应用实践能力和实践操作技能。通过座谈、观察、检测等方式观察学生的日常实践表现，了解学生的日常实践活动及进程。

(2) 社会实践总结。可采用教师评价或实践单位评价。主要检测学生的实践操作能力、知识掌握和组合归整能力、表达能力和组织协调能力。学生就参加社会实践的情况进行全面总结，包括实践方法、实践过程和相关实践经验，最后以实践报告或社会调查报告等方式反馈自己的社会实践成果。

## (四) 评价标准的弹性设置

形成性评价的评价标准应弹性设置。教师应针对每一门课程不同的教学内容和教学活动，设计相宜的评价方式。在实践操作中，建议各方式在总评价中所占的比如下：评价主体方面，教师评价占40%—60%，学生评价占20%—30%，实习单位评价占10%—30%；评价内容方面，命题笔试占30%—50%，日常学习表现占10%—30%，平时作业占10%—20%，专题论文占10%—20%，合作学习或讨论总结占10%—20%，其他占10%—20%。因为形成性评价主要是培养学生的学习自主性，增强

学生自主学习意识和学习能力，因此日常学习表现建议所占比例要相对高一些。

各方式的评价标准都要有相应规定。命题笔试、学习纪律方面如网络在线学习时间、作业上交次数等尽量执行量性评价，作业完成质量、专题论文质量、合作学习、讨论等建议质性评价。

### 三、形成性评价在高校课程评价中的应用

#### （一）建立评价学生全面发展的指标体系

形成性评价要为学生个性化发展提供依据与支持。评价指标体系包括学生的学科专业学习目标、一般性发展目标和个性化发展目标。其中学科专业学习目标主要是指学生经过专业学习掌握的专业理论知识和技能等方面；一般性发展目标主要是指学生经过专业学习所具有的多种学习能力及相关辅助能力，如自主学习能力、合作学习能力、思维辨析能力、信息获取能力、沟通交流能力、自我管理能力等方面；个性化发展目标是指学生根据自己的兴趣爱好和特长，强化自己某一方面或几方面的能力。

#### （二）实行形成性评价与终结性评价相结合

形成性评价并不是要全面摒弃终结性评价，在我国高校课程评价中，形成性评价应与终结性评价有效结合，既要包括以纸笔测验为主要方式的量的方法，也要包括以观察、谈话、收集各种反映学生学习过程与学习结果的作品在内的质的方法，既要注重过程的评定审核，也应重视阶段性测评和总结测评。教师要根据全方面的结果考量，为学生定量一个合理的测评结果，并通过这种测评结果实现学生能力的最优发展。

#### （三）所实施的形成性评价要能给教师带来情况反馈

形成性评价的根本目的是通过评价，及时了解学生的情况，发现教学过程中存在的问题，及时改进教学方法，修正教学过程，然后再评价，再修正，循环互动，建立“目标、过程、评价、反馈、修正”教学互进模式，真正实现以课程评价推动课程教学改革。对于教师来说，测评结果要能正确反馈自己的教学成果及学生对知识的理解能力和学生能力的发展状况等。因此，在形成性评价中，教师要认真做好考核评分工作。特别要重视考后分析工作，并及时调整教学内容和方法。

#### （四）所实施的形成性评价要能促进学生的自主学习能力

对于学生来说，测评结果要能及时发现存在的问题，在教师的引导下，通过

对自身学习计划的调整、学习方法的改进来达到提升学习能力的目的。形成性评价需要学生花费更多的时间与精力，需要学生积极主动地学习，需要学生主动配合教师，共同完成教学任务。

#### （五）加强形成性评价管理制度建设

高效规范的评价管理制度是保证高校课程评价改革顺利进行的前提条件，因此，高校要加强相应的评价管理制度建设。除常规评价工作管理外，还应包括评价反馈和改进机制以及课程评价改革奖励保障机制。

（本文摘自《现代教育管理》2011年第9期）

## 美国著名大学教学评价的内容特征

蔡 敏

**摘要：**教学评价是提升高等学校教育质量的重要手段，而评价内容的确定又是实施教学评价的关键性环节，直接关系到教学评价的指向和功能的发挥。美国著名大学在教学评价内容的选择上提供了很好的范例，体现出系统性、向生性、针对性和可测性等特征，对我国建设世界一流大学，推进教学评价的改革，有一定的参考价值。

**关键词：**美国；著名大学；教学评价；内容特征

高等学校的教学评价是一项由多个部分组成的系统性工作，包括评价制度的建立、方法和形式的选择、内容的设计以及结果的使用。其中，评价内容的确定最为重要，也最有争议。长期以来，对于“高校教师的教学评价究竟应该检查什么”，一直存在着不同的观点。由于教学评价内容能够直接反映一所大学的教育理念与教学要求，引导教学改革的方向和进程，所以探讨教学评价的内容问题必然成为我国高等学校面临的重要任务。为了在教学评价内容的研究中引入国际视角，对我国建设世界一流大学提供教学评价改革的参考和借鉴，笔者采用文献分析的方法，并结合本人在美国留学的经历，对美国10所著名大学的教学评价内容进行了全面的分析，总结出如下特征。

### 一、系统性：评价内容贯穿于整个教学过程

从教学系统论的观点来看，教学过程是多个有目的、有顺序的教学环节的集合，各环节之间有机配合、相互影响，共同对教学目标的最终实现发挥合力作用。因此，要想全面、客观地评价教师的教学质量，必须对教学中的每一个环节都进行认真的了解和检查。美国著名大学在教学评价的系统性方面为我们提供了很好的经验。从哈佛大学、麻省理工学院、普林斯顿大学、康乃尔大学等高校的教学评价工具中，可以清楚地看到，这些大学的教学评价内容涵盖了教学过程的各个环节。

#### （一）授课前的准备

教师在教学之前的周密设计和充分准备是取得教学成功的根本保证，也是教师良好教学态度的真实体现。美国大学教师在上课之前要做许多方面的准备工作，包括编写课程指南、选择学生参考书目、收集与研究教学案例、确定课堂讨论题目、挑选与复印学生阅读材料、安置和调试教学仪器、设计课后作业、开发测验

和考试工具等等。每一所著名大学的教学评价都非常重视这些方面，仔细检查教师是否认真完成了课前的准备工作。学校普遍采取的做法是，要求教师将上述反映教学准备情况的纪实性材料集中起来，交给校教师教学评价委员会，作为反映教学质量的基础档案或佐证材料。除此之外，大多数学校还让学生根据自己在学期中的观察，对教师的教学准备进行如实的评价。通常情况下，教师课前准备的权重约占总体教学评价的15%。

## （二）课堂教学的实施

美国著名大学对课堂教学有许多具体而独特的要求，这可以从各校学生评价教学的项目表中反映出来。课堂教学评价的内容标准可以归纳为如下10个方面：1) 用清晰、准确并且幽默的教学语言，讲述抽象的学科内容；2) 教学的内容和进度适合学生的知识基础和能力水平；3) 使用有效的教学方法，激发学生对专业学习的兴趣和渴望；4) 以分析和质疑的态度，介绍学科前沿理论以及发展趋势，不是一味地向学生传授陈旧的专业知识；5) 及时、科学地调整教学内容，使其与专业领域的变化密切相关；6) 鼓励全体学生积极参与课堂讨论，辨析和评判各种观点，分享各自的知识、经验和感悟；7) 认真、清楚地回答学生提出的问题，并且公平照顾到每个学生的学习需要；8) 给学生布置有价值的课后作业、研究项目和考试题目；9) 在教学中能够有效地利用课堂时间资源；10) 始终保持旺盛的教学热情和严谨的教学态度。

## （三）课余时间的辅导

除了全面评价教师在课堂上的教学行为以外，美国著名大学还特别注重考查教师在课后的教学表现。斯坦福大学、普林斯顿大学、加州大学柏克莱分校、波士顿大学等学校都对教师的课余辅导进行追踪检查。评价的问题涉及到：教师是否为学生安排了方便、足够的来访和答疑时间；是否及时批改和反馈学生的作业、论文或报告；是否及时回应来自学生的各种形式的课业询问，如电子邮件、电话、信函等；是否提供课外的实验技能辅导（理工科对此都有要求）；是否认真指导学生小组研究等等。通过对这些方面的评价，学校便可以清楚了解教师在课下是否向学生提供了全方位的帮助。

从上述评价内容可以看出，美国著名大学的教学评价关注的是教学的全过程，追求的是教育卓越。这与“教师教学评价就是对课堂教学的评价”的狭隘观念，形成了非常鲜明的反差。评价内容如果只涉及教师的课堂表现，不但不能全面反

映教学的真实情况,而且还容易产生误导,使教师错误地认为,“只要把课讲好,便圆满完成了教学任务,不必在其他方面花费更多的时间”。实际上,从大学教学的特点来看,许多重要的工作都需要教师在课下做,没有课前的精心筹划与准备,以及课后的及时讨论和辅导,不可能获得良好的教学效果。在大学教育中,师生之间的人际交往、思想碰撞和相互合作,是促进学生发展和专业成熟的最重要的条件。美国著名大学之所以能够培养出世界顶尖级的创造性人才,就是因为他们一直严格坚持“师生交流,共同探究”的教学原则,强调通过完整的教学过程来培养学生的学术思想和研究态度。

## 二、向生性: 评价内容关注学生学业收获

所谓教学评价内容的向生性,是指评价项目不但检查教师的教学态度和行为,而且充分了解学生通过教学在各个方面所取得的进步。另外,向生性还体现在,评价不仅关注教师对课程体系和知识内容的处理,而且注重教师在教学中对学生的注意、尊重、关心和帮助。在笔者所研究的美国著名大学中,绝大多数学校都把学生对于教学的亲身感受作为教师教学考核的重要内容,不单纯评价“教师做得如何”。这种评价取向明显反映出美国大学“以学评教”的理念,试图通过对学生学习情况的检查,来衡量教师教学的有效性。一种在美国大学里广泛持有的评价观点是,只有学生在课程学习中取得了多方面的明显进步,才能判定教师的教学是成功的。学生在如下几方面的收获,特别受到教学评价的关注。

1. 学科理论知识的理解和掌握。例如,哈佛大学在学生评教中问到:“教师的授课对你理解学科内容有帮助吗?”“教学过程是否使你提高了分析能力?”波士顿大学也问到类似的问题,例如,“教师对问题的解释是否有助于你深刻理解课程的内容?”“教师所使用的教学方式对你掌握该学科知识是否合适?”“课堂的讨论能否提高你的解决问题的能力?”纽约州立大学则向学生提出了这样一些问题:“学习这门课程以后,你的学科知识增加了许多吗?”“你是否能够将学到的概念和原理应用于相关的问题情境?”所有这些评价内容,都是在考查教学对学生的知识学习和能力发展的促进作用,是对教学质量的直接检验。

2. 学习课程的态度、兴趣和愿望。加州理工学院在评价教学中就学生的学习态度这样问到:“你愿意上这门课吗?”“你希望在课后与教师一起讨论问题吗?”康奈尔大学针对学习兴趣向学生提出如下问题:“教师的反馈能否激发你进一步探究学科知识的兴趣?”“教师设置的问题是否能把你带进研究的乐趣之中?”

普林斯顿大学还问到了学生的学习愿望：“你是否渴望获得该课程更多的专业知识？”“课程是否激发了你在课下探讨问题的愿望？”“教学是否激励你建立更高的学习目标？”从这些评价问题可以看出，美国著名大学的教学评价非常重视学生的学习情感反应，把学生在学习过程中的心理情绪作为判断教师教学效果的重要指标。

3. 学习的独立性和创造性。美国著名大学教学理念的精华是提高学生独立学习和独立创造的能力，把学生培养成为具有独到见解的开拓型人才。这一目标在每所学校的教学评价中都呈现得非常清楚。斯坦福大学在教学评价中问学生：“教师的讲授能否使你对教学内容产生新的想法？”“教学内容是否可以引起你的创造灵感？”“你在课程学习中是否敢于提出自己的问题？”加州大学伯克莱分校在教师职称晋升的评价中，详细考查教师在培养学生的独立性、好奇心、推理性等方面的做法，检查他们是否培养新生的独立学习习惯，是否向优秀学生布置具有挑战性的学习任务。

美国著名大学在教学评价内容中体现向生性的做法，是我国大学应该借鉴的。笔者对国内许多大学的教学评价工具做过统计和对比，发现评价内容主要集中在教师的教学表现，极少提及学生在教学中的进步与变化。即便有些学校的教学评价考虑到了学生的学习收获，但是还非常笼统、粗糙，没有展开细致的了解和评价。就连目前各高校普遍参照的《精品课程评审指标体系》，对学生在课程中“学得怎样”也关注的很少。该指标体系虽然有30多个观察点，但仔细分析后可以看出，仅有2条与学生的学习行为有关。如果教学评价不考查学生的学业成就，只看教师是否达到了预定的标准，并不能保证教学评价结论的全面性，也无法客观判断教学质量的高低。因为真正的教学效果取决于学生方面的变化，尤其对于大学教学来说，学生在课程中的表现更应当引起教学评价的关注，只检查教师单方面的努力是远远不够的。

### 三、针对性：评价内容体现学科课程特点

美国大学教学评价显现出很强的专业针对性，不同学科有不同的评价内容，充分体现了高校教学的多样化特点。笔者从本次研究中了解到，美国著名大学通常采取的程序是，首先在学校层面制定教学评价的一般性指南，确定出对教师教学的基本要求，然后各个院（系）再根据自身专业的特点，编制出符合学科教学实际的评价工具。这样，评价的内容既能反映一所大学教学的共性要求，也能体

现不同专业的个性特征，有效避免了“一表多课”的评价弊端。哈佛大学、麻省理工学院、加州理工学院、斯坦福大学等学校在这方面做得尤为突出，不仅每个学院的评价表格不同，而且其中每一个专业的评价项目也不尽相同。这些学校的评价表上都标有课程的编号和种类（如理论课、实验课、技能训练课等），以便指示出所适用的专业课类型。除了体现专业特点以外，教学评价还考虑到了教师的类别，比如，在评价实习教师、实验教师、助教等人员的教学情况时，所使用的评价表就与正式授课教师的表格不同。这种个性化的评价，可以将各类课程的教学特征反映出来，对鼓励每个教师形成自己的教学风格，发挥个人的创造潜能，具有很大的激励作用。

#### 四、可测性：评价内容指向具体教学行为

教学评价是一种以实事判断为基础的价值判断活动，教师在教学中的具体表现和学生在学习中的实际进步，是获得评价结论的实事依据。没有对教师行为和学生行为的真实描述和充分了解，不可能对教学质量做出正确的评判。因此，在进行教学评价时，应当选择那些可以观察到的教学事实作为评价素材，保证教学评价项目的可观察性。美国著名大学的教学评价内容很好地反映了这个要求，所采用的教学评价标准具有明确的观察性含义。例如，是否允许学生发表不同或相反的学术观点、是否有效激励和引导课堂讨论、教师的教学是否能够激发学生课后阅读专业书籍的兴趣、教师给出的论文评语是否对学生有具体的帮助、教师提供或介绍的课程材料是否完整、相关和领先等等。对于教师在这些方面做得怎样，学生完全可以根据自己的切身观察，做出客观、准确的评价。相比之下，目前我国大学教学评价的内容比较缺乏可观测性，许多评价题目不是引导评价者注意教学事实，而是让他们对较为抽象的概念给出模糊的等级评价。例如，“教学观念”、“教学能力”、“教学风格”、“教学态度”、“教书育人”等，经常出现在学生评教的表格当中。由于这些教学“术语”不能直接反映教师的具体行为，只是一个相对笼统的概括，所以许多学生就凭着自己的“感觉”，给教师打一个印象分。从这一点看，美国大学对教学评价内容的确定要比我们务实得多，评价项目显示出很强的操作性，对教师教学表现的评价更符合真实情况。我国大学应当在这方面多做一些努力，让教学评价的项目与教师的教学行为一一对应起来，使评价的结果具有明确的含义，从而能够为教师日后改进教学提供具体的指导信息。

（本文摘自《外国教育研究》2006年第6期）

## 影响评估效果的因素

唐纳德·R·克里克山克 德博拉·L·贝勒尔 金·K·梅特卡夫

为了进行有效的评估,教师们必须考虑到一些影响他们评估准确性和有效性的因素。首先,他们必须选择适当的信息源;其次,他们必须确保他们的测评和判断的准确性;最后,他们必须知道他们将怎样利用这些评估及其利用的意图所在。

### 信息的来源

在第四章中我们讨论过一些信息的来源,如积累的记录、观察、访谈等一些有助于老师更好地了解自己的学生信息。这些信息源让老师更清楚地把握学生的优缺点、好恶、存在的问题和取得的进步。教师利用这些信息组织课堂教学和制定课程计划。在这一章里,我们主要讨论这样的信息源:这些信息源有助于确定哪些知识学生已经掌握、哪些方面或哪些方法还需要进一步施教。虽然第四章中讨论过的一些更多的信息源也能用来评估教学成绩,但这类信息多数是教师通过分析学生的作业或测验而获得的。

### 对学生作业的分析

显然,关于教学成绩的大量有用的信息是通过分析学生的作业而获得的。在学生的书面作业比如所布置的家庭作业中,老师不仅仅只注意一个学生所得的分数,他们经常想知道该学生是如何得到一个答案的。例如,这孩子解题的过程和应用了什么方法,他或她使用该方法的准确性如何?通过仔细地察看学生的解题过程,老师便能够确定,是什么样的错误导致了该学生未能准确地完成作业。在布置的家庭作业、练习、论文以及各种大大小小的测验中老师都能对学生的学习进行观察和分析。当老师做此工作时,说明他们对学生所完成的作业给予了反馈信息,这一点非常重要。

### 测验

测验是最被公认的评估信息源。测验要求学生展示他们对某一确定部分知识内容的掌握程度,从而使得老师得以对学生的学习成绩进行表述。测验与其他一些教育活动一样,还能用来让学生被迫展示他们是怎样解答问题或怎样得出结论的。因此,测验是一种丰富的信息源。然而,跟其他信息一样,测验所提供的信

息只有在确保它的准确性和公正性的情况下才是有用的。此外，测验只是学生学习进步程度的一个尺度。

有效的评估要求教师对尽可能多的信息源产生的尽可能准确的信息进行收集和综合。他们必须将他们对学生进行非正式观察和交流中获得的素材与他们通过测验、各种学生作业及其他可能的源泉所获得的更加正式的测评结合起来进行判断。然而，老师的这些判断的价值很大程度上取决于他们所获得的信息的准确性。

### 信息的准确性

教师所获得的有关学生及其进步的资料不可能是完全准确的。基于种种原因，没有哪一种测评方式能够完全避免偏见或错误。事实上，刚刚提到的一些信息源就明显带有不准确性的特点。例如，教师或家长的看法就很容易倾向于某种程度上的不准确性，因为他们的看法具有高度的主观性。同样地，教师对学生的观察也不可避免地具有选择性和不完全性；它们会受教师们的个人价值观影响从而导致错误和偏见。因此，在收集和运用源于这些活其他源泉的信息时，教师必须留心警惕一些潜在的偏见和错误。

甚至测验也易于导致错误。没有一个测验能够对一个学生在某学科领域的知识和成就进行精确的测量。首先，测验所考察的内容是有选择性的。它们通常只能检测一个教师自己力图教授的内容的一部分。同时，它们无法检测学生从书本、朋友或其他教师等其他途径所获得的知识。它们也可能忽略一个学生从出题教师那里所学到的一些偶然性的知识。而且，该学生在测验时的内部状况也可能影响考试结果。也许，该学生在考试时感到疲倦或身体不适。也可能测验中所强调的内容碰巧被该学生临考前复习到了。在上述每一种情况下，该学生所得的分数都不是他对这一学科所掌握的知识的全面或准确的测量。

一位教师在确定通过测验所获得的信息的准确性时主要应该考虑如下主要因素：（1）该测验是否检测了它所意欲检测的内容（效度）；（2）该测验中是否体现了一贯的测量标准（信度）。基于效度和信度这两个因素与测验密切相关，在判断从其他源泉收集的信息时也同等重要，因此我们将对这两个因素进行探讨。

### 效度

效度（validity）这个术语是指一个测验相对于它意欲检测的内容而言所检测到的程度。一个测验的效度依据以下三个标准而定。第一，一个有效的测验应

该恰到好处地检测试剂所教授的内容。换句话说，该测验应尽可能多地囊括整个教育过程所覆盖的内容，但同时，它又不能含有不要求学生掌握的内容。特别地说，一次有效的测验能检测出学生对老师所讲的每一个主要的知识点的掌握情况。

第二，一个有效的测验应该测量出教师教育在各个领域（认知、情感、动作技能领域）所达到的各种水准（见第五章布鲁姆的分类法）。例如，如果一位历史老师首先主要讲授了有关美国革命的知识（史实），然后花大量的时间让其学生对这些史实性的知识进行分析、综合、评价，那么，随后的有效的测验就应把重点放在分析、综合、评价这三种高于史实的认知水准上。

第三，一个有效的测验除了应该反映出教师教育在各个领域和不同水准之间的平衡以外，它还应突出反映教师所教的各方面知识的重点所在。例如，如果在传授有关美国殖民地时期的意识形态的过程中，关于该时期的社会条件和经济条件的知识内容所占据的教育时间大体相当，那么，随后的有效的测验中这两方面的内容所占有的比例就应基本相等。

也许，再举一个例子是有用的。让我们假设你讲授一个有关奴隶制度的单元内容，为时 3 个星期，你的讲授基于 4 个认知目标——一个基于知识水准、两个基于综合水准、一个基于评价水准。在讲授过程中，你用一天时间讲解美国奴隶制形成的原因，3 天讲授奴隶制度的盛行，3 天分析促使奴隶制衰亡的因素，两天讲解废奴宣言对奴隶制度和南方经济的冲击。一个有效的测验应该要求你的学生对原因、盛行、衰亡的因素和冲击这几个主题的理解进行展示。此外，该测验应该让学生运用他们所学的知识进行以下事项：（1）回忆一些信息（知识水准的目标）。（2）综合一些信息。（3）对一种或多种情形进行评价。该测验的首要重点应放在有关奴隶制度盛行和衰亡的主题上（因为每个主题占据了 3 天时间），次重点应放在花费了两天时间的废奴宣言的冲击上，最后才是只花了一天时间的奴隶制度形成的原因。学生不应被问及有关其他主题的问题，如有关奴隶制度对非洲文化的冲击的问题，即使这些问题有可能在这个领域里被偶然地学到过。

总而言之，效度是指一个测验检测你预期它检测内容的准确性。因此，它能为你提供有关学生的学习情况和你的教育效果的信息。通过我们的讨论你也许已经明白，任何一个测验都不可能完全符合效度的每一个标准。效度不是一个二选一的问题，而是存在着一个程度的问题，因此，对于一个特定的测验来说，你必须考虑它对你的目的是否足够有效。在随后的章节中我们将讨论一些如何让你的

测验更有效的方法。

### 信度

除了检测你所期望检测的内容以外,你也许还希望你的测验能够提供测量学生学习情况的一贯的测量尺度。这种一贯性就是指信度 (reliability)。特别地说,一个测验的信度意味着,如果让学生多次进行该测试,每次所产生的分数应该是几乎近似的。而事实上,没有一个测验能在学生进行该测验的不同次数中都提供完全一致的分数,因此,与效度一样,信度也必须被看做是存在着一个程度的问题。

由于测验本身的某一错误,一个缺乏信度的测验所提供的分数可能不能准确地反映学生的实际水平和能力。也许可用一个类推的例子来帮助理解。让我们假设你需要测量你教室里书架的长度。你先用一根从窗台上取来、被强烈的太阳光照射过的、细细的金属量尺来测量书架。你记录下 4 个书架的长度。然后你被办公室里的一个电话给打断了。你把这个量尺随手放在空调的排气管上,就去办公室接电话去了。当你从办公室回来时,你意识到你把记有测量数据的纸忘在了办公室里。你不愿回到办公室去取,便从空调排气管上拿起量尺,重新测量了刚刚测量过的 4 个书架的长度。当天晚些时候,你发现了忘在办公室里的那张纸,并注意到,对这 4 个书架所进行的两次测量的数据不一致:第一次测得的书架的长度比第二次的略微偏小。你有可能作出结论说,暑假的长度在那短短的时间里起了变化,然而,正确的解释却是:这个金属尺没能为你提供可靠的测量,因为它在被置于窗台上的阳光下发生了热膨胀,从而导致量得的书架长度显得偏小,而在被置于空调的排气管上以后发生了冷收缩,从而导致量得书架的长度显得偏大。

当一个测验不具信度时,也会出现同样的问题。缺乏信度的测验会导致我们对学生的学习作出不准确的结论。如果学生们所参加的测验本身有缺陷,不能提供可靠的结论,那么,有可能,当学生们事实上已经掌握得很不错时,通过测验却显示出他们所取得的进步很小。信度是很难评估得。一些商业性用途测验的制定者通常为它们的测验提供一些信度的统计指标,这种指标被称为信度系数。信度系数从 0.00 到 1.00 排列,0 代表不具信度,1.00 表示完全的信度。一般来说,信度系数如果高于 0.85 就被认为是可以接受的。尽管你也许不能计算出你自己的测验的信度系数,但是你可以通过增加测验所包含的问题数、通过确保评分的客观性和一贯性来提高你的测验的信度指数。在随后的章节中将介绍建构信度

测验的指导方针。

### 评价的类型：形成性评价和总结性评价

我们已经明白教育评价所运用的信息来源众多，而且，这些信息可能具有不同程度的准确性。教育评价的用途也各种各样。一般来说，我们认为，教育评价时出现在教育行为之后用以判定学生分数的。然而，斯克里文 (Scriven) (1967 年)指出，教育行为之后的评价对于提高教学并不是很有用。虽然评分是必要的，但教育行为之后的评价并不能帮助教师或学生纠正其不足之处。斯克里文认为，教育行为过程中的评价更为重要。在教育行为进行过程中，这种评价有利于教师调整其教育方法，以便适应学生的需要。

斯克里文介绍了两个术语来描述这些不同的评价类型，一个是形成性评价，一个是总结性评价。形成性评价是指教育行为过程中的评价。这种评价为教学提供反馈信息以便及时可能地影响教学过程。虽然形成性评价也可以对学生的成绩进行评分，但其主要目的是提供反馈信息，以有利于制定和改变教育计划。形成性评价能够促使教师采取有效的教育措施，从而提高学生的学习成绩。形成性评价既包括教师所进行的非正式的观察，也包括学生的各种表现事例，如教师提问的反应，或在一个较大的教学单元里学生所完成的任何作业。通过形成性评价所获得的反馈信息能帮助教师在教学的总结评价之前调整教学方法，提高学生成绩。

总结性评价是指教育行为完成之后所作出的评价。这种评价是用来对学生的学学习进行最终的判断。其主要目的不是调整教育方法或纠正学习中的不足，而是对学生的成绩和进步进行总结，一般采用等级或分数的形式。总结性评价一般包括从单元测验、期末考试或其他学期末的考试项目中收集的信息。多数情况下，在这种总结性评价之前就已经进行了相似的、但所反映学生表现更少的形成性评价。

优秀的教师常常运用形成性和总结性两种评价。然而，通过各种各样正式和非正式方法形成并随后传达给学生的形成性评价才是最有利于提高教学质量 (Brophy, 1981)。表 8-1 对形成性评价和总结性评价进行了大致的比较。

表 8-1 形成性评价和总结性评价的特征

	形成性评价	总结性评价
目的	在教学过程中监控和指导教学过程	教学过程结束时对教学效果进行评判
评价的时间	教学过程中	教学过程结束时

评价技术类型	非正式的观察、小型测验、家庭作业、学生的问题、学生的作业单	正式测验、作业、试卷
评价信息的运用	在教学过程中提高和改善教学进程	对教学的总体效果进行评判；评分、排列名次、并予以鼓励

(本文摘自《教学行为指导》第八章)

## 教师自行设计的考试

唐纳德 · R · 克里克山克    德博拉 · L · 贝勒尔    金 · K · 梅特卡夫

标准化考试对于区、州以及国家机关制定教育政策起着日益重要的作用。然而,对于班级评价来说最普通的形式仍然是教师自行设计的考试(Borich, 1992)。据估计学生们将自己课堂时间的 10%以上花在了配合教师的评价行为上(Crooks, 1988)。教师自行设计的好的测验相对于标准化考试来说有如下几个优点。第一,既然是教师自己出的考题,这些考题的内容就必然能着重于该教师所讲过的自认为较重要的内容和知识点,而不是像标准化考试那样对大多数老师在该阶段可能讲授的总体内容进行笼统的检测。第二,教师自行设计的考试远比标准化的考试节省费用。第三,由于这种考试的内容紧紧围绕所讲的知识点,因此,它能提供更为详细具体的信息,以便让教师更为准确地知道学生对所学部分知识内容掌握得怎样。虽然教师自行设计的考试有如此多的优点,然而,这种考试经常因为设计得不好而导致它所提供的信息不够准确(Haertel, 1986)。

鉴于教师自行设计的考试是如此普遍并具有如此大的潜在价值,我们现在就进一步讨论一下如何更好地设计并更充分地利用最原始的两种教师自行设计的考试:一种是纸笔测试,一种是学生行为考评。首先我们将对这两种考试类型的每一种加以定义和说明,然后讨论其优缺点,指出在选择和设计合适的考试时应该考虑哪些因素,并提供一些评分的方法。

### 纸笔测试

教师进行自行测评时最易被认可的一种方式就死纸笔测试。我们每个人都经历了不知多少小时的纸笔测试,要么是对一些评论性问题长篇累牍地发表自己的看法,要么是在一页页的答题卡上用铅笔没完没了地将一个个小圆圈涂黑。此外,也许常有此类情况发生:要么是考试中有的问题提得模棱两可,含糊不清,甚至下了考场后绞尽脑汁也弄不明白老师究竟想问什么;要么就是考试之后我们百思不得其解,为什么老师就某一个知识点反复出了几个问题,而对有的知识点却全然疏忽。不管上述哪种情况,都会因为考试试题设计得不好而无法真实地反映我们对所学知识的掌握情况。作为一名教师,确保你的学生有机会真实地展示他们

对所学知识的掌握情况便是你的应尽职责。

纸笔测试通常采用两种最常用的题型。一种题型是要求学生从众多可能的备选答案中选出一个正确的答案，通常被称为客观性选择题。另一种题型是要求学生用自己的语言进行阐述或评议。这种题型被称为主观性试题。评价题或简要回答题都属于主观性试题。

就出题和评分的难易程度来讲，每一种题型都各有优缺点。换句话说，其中一种题型出题更容易，但评分更难，而另一种题型恰恰相反。不少人总是喜欢过去绝对化地认为其中一种题型毫无条件地优越于另一种题型，比如，有人认为评价题比选择题更有价值。然而，摩尔（Moore）（1992 年）写道：

“一个普遍存在的错误观点认为，选择题的考试仅仅针对一些死的知识点。事实上，任何类型的考试都能用于检测任何水平层次的任何教学成果。虽然要出好高水平的选择题总是非常困难，但是，要想出好任何一种题型的高水平试题都绝非轻而易举之事。”

因此，我们应该弄清楚并学会利用每一种合适的题型。

### 主观性试题

主观性试题要求学生基本上用他们自己的语言作答，而不是从一组答案中选择一个合适的答案。主观性试题能够要求学生从短短的一句话展开讨论，发挥想象，进而写成一篇数页的议论文。设计得好的主观性试题能够使学生得以展示出所有认知层次的知识——从回忆书本知识到进行主观评价。好的主观性试题能促使学生对一些零碎的知识点和概念进行组织或再组织，并用各自不同的方式对所学内容进行阐释，进而解决问题。设计得不好的主观性试题可能无法激励学生的创造性，混淆学生的概念，模糊学生的思路，从而导致对学生的学习进步情况做出不公正、不准确的判断。

主观性试题有如下几个优点。第一，出题容易。的确如此，设计主观性试题远比设计选择题容易得多。当教师需要对学生的高水平的认知过程进行评价时，这一点显得尤其正确。因此我们说，主观性试题所花费的出题时间较少。第二，由于学生在答题时必须用自己的语言进行组织和表达，主观性试题使得教师有机会对学生的语言表达能力进行评价。第三，主观性试题能够减少学生纯粹猜测答题的可能性。

主观性试题也有缺点。虽然设计主观性试题既容易又快捷，但主观性试题的

阅卷评分工作特别费时。例如，假设一位高中老师要给 4 个班的学生进行一次主观性试题考试，每班 30 人，每份试卷 5 道题，即使每道题学生只答半页，这位老师也将不得不批阅 300 页的学生答卷！要对学生在这次考试中表现进行及时而详细的反馈恐怕是件令人望而生畏的事情。

主观性试题的另一个缺点是考试中所能囊括的内容总量比较有限。因为必须给予学生足够多的时间分析题目、构思答案内容、并将其构思表述出来，所以一次考试只能包括几道题目。因此，一次完全由主观性试题组成的考试不可能对学生的全部知识进行检测。实质上，既然无法涵盖所有重要的知识点，那么，该考试的效度必然大打折扣。

主观性试题考试的最后一个、同时也是最麻烦的一个缺点是很难对其进行可靠的评分。主观性考试要求教师对学生回答问题的质量进行判断。有好几个因素都可能使这种判断产生偏差，从而导致评分的不公平性。例如，学生的书写是否工整清晰、是否有语法错误、写作风格如何、答题的篇幅多少等因素都会影响教师对主观性试题的评分 (Stanley, 1958)。这一点对于年幼的学生或者不够精通英语的学生来说尤其可能是一个问题。另外，教师对每一个学生的固定的看法和对他们考试成绩的预先期望都可能导致他们评分不公正。例如，一个平时课堂讨论积极、主动、热情的学生就很可能比一个平时课堂表现不主动积极、而考试中答题情况一样的学生更容易获得高分。

主观性试题可分为限制性和非限制性主观试题。限制性主观试题为学生答题规定了明确的限制条件。这种试题可能限制该题所包含的主题，也可能限制学生的答题篇幅，此外，限制性主观试题还经常建议学生运用哪些知识点进行作答。限制性主观试题无法让学生充分发挥其灵活性及创造性。然而，正因为限制性主观试题对学生的答题风格、篇幅和主题都减少了可变性，从而使得它们从总体上讲要比非限制性主观试题更具信度和公正性。

非限制性主观试题通常对学生的主观反映不做任何限制。因此，非限制性主观试题能让学生发挥出更大的创造性和灵活性。学生答题的篇幅可随他们的意愿而定，可多可少。而答题的内容可以包括他们认为必要的所有知识点。非限制性主观试题能让我们更好地把握学生是如何表述并创造性地解决问题的，但同时，对这种试题的评分更容易趋于偏颇和不公正。基于此，你还是应该在大多数情况下才用限制性主观试题。限制性和非限制性试题的例子请见图 8-2

**限制性主观试题**

列举并简要阐述波 (Poe) 在“招待员之家的秋天” (“The Fall of the House of Usher”) 一文的首段中运用了哪三种文学手段从而确定了全文的基调，并分别举例说明。所答内容不得多于五个句子。

**非限制性主观试题**

对波在“招待员之家的秋天”一文的首段中所运用的确定全文基调的文学手段举例说明并加以论证。

**图 8-2 主观性试题：限制性和非限制性主观试题**

既然你已经意识到你最好采用主观性试题来检测你所讲授过的知识点，你可以采取如下几项措施以提高主观性试题测验的效度和信度，从而提高测验的准确度。第一，让测验内容紧扣所学的知识点。你对学生进行测验的目的是创造教学评价工具以便提供准确的尺度，用以衡量出每个学生对所有知识的掌握情况。因而，该测验应该在广度、深度上都具有代表性，突出知识重点，并能促使学生结合你所教的知识点进行作答，同时展示出自己的认知水平。只要你能先把所教的知识点都清楚完整地写下来，那么，你会发现，设计一套主观性试题测验将是一件相对容易的事情，请遵照以下指导原则。

第一，确定你所需要检测的最重要的知识点是什么。先前我们已经指出，学生解答主观性试题比较费时，因此，你可能无法对所有知识点进行检测。在这种情况下，你应该挑选出学生所掌握的知识材料中最关键的、教学过程中最强调的、或者那些涵盖面较广、包含性较强的知识点。但同时，你还必须很现实地考虑主观性试题的题量，不能超出学生的能力范围。所有的学生都应该在允许的时间内能够答完所有的题。

第二，针对每一个检测目标设计出主观性试题。通常从动词部分开始(比较、对比、列举等等)，并提出限制性条件，如所要求的答题篇幅、所运用的知识要点以及所采用的文体风格。最好还能指出拼写和语法错误是否纳入评分因素之列，是否允许拟出提纲，以及每道题的每部分所占的分数比重。

第三，出完题之后，你最好马上作出一份完整的参考答案出来，以便对你的评分工作起到指导作用。记录下来你所花的答题时间，并以此决定学生完成该测验需要多少时间。(请记住，学生要比你花费多得多的时间去构思、组织并表述出他们自己的答案。) 在做好满意的参考答案之后，你应将答案所包含的各个要

点进行细分、罗列，以形成评分时所需的模板。

第四，要尽量增加题目数，并对每道题要求相对较少的答题篇幅，而不是相反。这样能使你检测到更多的知识点（提高测验的效度），并能给学生一个更好的机会准确地展示自己所学到的内容（更具信度）。此外，让所有学生对所有的考题进行回答而不是选作其中一部分，也算数明智之举。

在将出好的试卷送去复印之前，你应该写出一份完成该测验的指示，其中包括：每道题的答题要点，你的评分依据和原则，以及每道题的每个要点所占的分数比重。在适当的时候也可以建议学生每道题所用的时间量。然后认真地通读试题，也许应该请一个同事来读，以便你听得更清楚，并请他参考一下你得时间安排是否合理。

正如前面已经指出，主观性试题的一个主要缺点是很难对其进行可靠而客观的评分。相对于非限制性主观性试题来说，采用限制性主观试题更有助于评分，因为限制性主观试题明确地规定了学生在答题中所应包括的信息范围和知识点。除此之外，如果你对每个学生的试卷进行审阅评分时都能参照你事先拟好的参考答案的各项要点，那么，你的评分会更具一致性和客观性。如何才能提高你评阅主观性试题的水平呢？指导原则如下。

第一，尽量做到在不知道你所阅试卷是哪位学生所答的情况下进行评分。你可以让你的学生将其姓名写于试卷的末页，或给每个学生一个代码以替代其姓名，或者将所有学生的姓名盖住，直至评阅完所有试卷。

第二，同时评阅所有试卷的同一道题。这样做会促使你对所有学生的答题内容更加一贯地采用同一标准。同时还有助于减少你审阅了一个学生的前一道题时对其答案的好坏看法对他后面所答题目产生的影响。

第三，不要急于一下子评阅完所有试卷。对主观性试题的评分工作是一件很费时间和精力的事情。如果你已经感到疲劳，你将更难与确保你评分的客观性和准确性。最好一步一步地先对所有学生的某一道题进行评阅，然后评阅下一道题，以此类推，直到你感觉有点疲倦时就停止下来。千万不要搞疲劳战术，试图长坐数小时以指望一次性攻克万堆如山的试卷。

第四，评分之后，重新浏览所有试卷或者随手抽出一份试卷加以审阅，以确保你评分的准确性。通常认为，最好将试卷先归类为高分、中分和低分三等，然后从每一类中抽出一部分试卷重新审阅，以检验你的评分是否准确。

第五,必要时,应该调整评分标准以适应因教育失败而导致的普遍水平过低。如果没有一个学生能够比较准确地回答某一道题,你可以考虑取消这道题,或者重新考虑这道题的评分标准。

主观性试题对于进行班级评价非常有用。由于主观性试题出题容易,不至于让教师在临考前感觉到时间过于紧迫。主观性试题也比较适合于检测学生的更高层次认知水平。然而,主观性试题的评分工作特别费时,而且很容易受到评分偏差的影响。这些特征致使一些人建议,只有当选择题无法被采用时才采用主观性试题(Borich, 1992)。表 8-3 概括性地介绍了主观性试题的出题和评分指导原则。

表 8-3 主观性试题的出题和评分的指导原则

出题的指导原则	评分的指导原则
1. 所有考题紧扣所学知识点	1, 出好每道题后立即拟定好参考答案
2. 选出最重要的知识点并对其中每个知识点出一道题	2. 标出你所拟参考答案的各个要点
3. 出题时首先考虑动词	3. 依照参考答案对所有学生的试卷评分
4. 尽量出限制性主观试题,而不是非限制性主观试题	4. 在不知道所阅试卷时哪位学生所答的情况下进行评分
5. 在学生可能完成的情况下尽量增大题量	5. 同时对所有学生的同一道题评分
6. 指出拼写和语法错误是否纳入评分考虑范畴	6. 避免在疲劳时阅卷
7. 在将试题拿去复印之前仔细通读一遍	7, 从评阅后的试卷中抽样检查你评分的准确性

### 客观性试题

客观性试题要求学生从几个备选答案中选出一个最合适的答案或者用一简短的答案(通常是一个单词或短语)将句子补充完整。学生不需要像在主观性试题中那样自行构思、组织各自的答案。前面已经提到,许多人认为,客观性试题不利于对学生更深层次的认知能力进行测评。但是,如果出题人认真细致,客观性试题也能够像主观性试题一样有效地激励学生展示深层次的思考能力。比如,

可以在题目中要求学生对一些隐喻手法进行识别，或者提出一些针对电影剧本的剧情说明、情节、外延涵义的问题。

客观性试题有许多优点。也许最为重要的一点就是其评分工作更具客观性——因为学生要么选对、要么选错了正确答案。评分的信度和公正性得以提高，因为分数不依评分者的判断为依据。而且，由于客观性试题花费学生的答题时间较少，因此，在客观性试题的考试中，教师可以对更多的教学内容和知识点进行检测。从而使得这种测验更具效度。评分工作简单、快捷、而且更为客观。由于客观性试题不要求学生自己构思组织答案，因而使得这种考试更能检测出学生对所学知识内容的掌握情况，而不是像主观试题那样检测学生的书写表述能力。

但另一方面，客观性试题的评分工作比较困难而且费时。因而临考前出客观性试题远比出主观性试题更容易遭受时间紧迫感的折磨。客观性试题的一个主要缺点就是，很难设计出能够激励学生展示出更深层次的认知能力的试题。虽然设计出这样的试题是能够做到的，但通常让人绞尽脑汁。正因如此，客观性试题通常强调的是浅层次的思考。

客观性试题有四种题型：选择题、搭配题、正误判断题和填空题。接下来我们将讨论客观性试题的出题、运用和评分工作。

我们大多数人在学习期间都已经参加了无数的选择题考试。不仅几乎所有的标准化考试都会采用选择题型，许多教师自行设计的考试中也采用该题型。一道选择题包括两个部分：陈述或者提出问题的题干和几个备选项。学生们得从这些备选项中选出题干中所陈述或提出的问题的最佳答案。

选择题有几个优点。选择题可能是客观性试题中用得最广泛的题型，它能用来对任何学习内容和所有认知层次的知识进行检测。由于学生完成每个选择题所花的时间相对于主观性试题要少，因而，教师能用选择题对所学的更多的知识点进行检测。或者说，选择题能使得教师更加准确地区别各个学生对所学知识掌握的深度和广度。相对于主观性试题来说，对选择题的评分容易可靠。

选择题有两大主要缺点。最大的缺点是出题困难而且费时，尤其是当你想设计能够测评出学生如何运用所学知识进行深度思考的试题时。要设计出高水平的选择题，首先要求就是掌握更高水平的信息，然后从该信息中提炼出一个问题，这个问题很可能诱使学生产生类似的理解，之后设计出几个比较具有迷惑性的备选答案，而每个备选答案都需要学生进行深层次思考。第二个缺点是，选择题无

法像主观性试题那样使我们得以评价出学生对所学知识的表述能力怎样。学生只是选择答案而已。

你可以采取一些措施来提高选择题考试的效度和信度。首先，与其他任何考试一样，在出题过程中，你应该使你所出的选择题与你的教学内容相吻合。你可以以教学大纲为指导依据，大体确定出每部分教学内容应该出多少道题。然后你就可以依照这个大致的考试计划而进行选材和出题。

选择题的出题过程从设计题干开始。题干必须陈述得简洁明了，而且应该提出或者陈述一个问题——而不能是一个不完整的陈述。题干应该表述得足够清楚以至于不看备选项也能让人确定出最合适答案。要确保每一个题干都提出一个新的问题，而且所有的题干和备选项之间不能有相互暗示的情况。在题干和备选项中都避免使用否定的问题或陈述，这也可谓明智之举。

例如，“如果英国军队没有\_\_\_\_\_，美国革命就可能不能取胜。”这样的题干本身就把人的脑子给搞糊涂了。它包含太多的否定，而又没有提出一个明确的问题。这个题干最好改写为：“英国军队所犯的下列错误中，哪一项是导致美国革命成功的关键性因素？”这个题干就避免了否定句式，提出了一个完整的问题，而且表述清晰。

写好题干之后，你就设计出 3-5 个备选项，将正确的选项列在首位，不一定要对每一个题干都设计出相同数量的备选项。为了减少学生猜测作答的可能性，每个备选项都应是对题干所提问题的一个看似合理的答案，但只有一个是最准确地。避免使用看似不合理的备选项，避免使用“以上各项都不”或“以上各项都”之类的字样，因为这样的备选项有利于学生猜测作答。确保每个题干及其备选项都不要对该题的正确选项或别的题干提供语法上的提示。当题干是一个不完整的陈述时，此类情况常常发生。例如，题干可能问：“George Washington was a \_\_\_\_\_”（“乔治华盛顿是\_\_\_\_\_”），这个例子中，根据语法知识就知道，正确的答案选项一定以辅音字母开始，如果“affluent land owner”（“富裕的土地主”）为其中一个备选项，根据语法规则就可判定它不应该是正确答案。

最后，每个备选项的长度应该大致一样（在设计得较差的选择性试题中，最长的备选项常常是正确答案）。图 8-3 对如何设计出好的选择题提供了概括性指导原则。

1. 写出题干，题干应该提出一个完整的问题。避免使用陈述不完整的题。
2. 题干的表述要明确清楚，即使学生不看下面的备选项也能知道其正确答案。
3. 在题干和备选项中避免使用否定的问题和陈述。
4. 避免在题干中为正确答案提供语法上的线索。
5. 首先写出正确的备选答案，然后想出 2 至 4 个合乎情理的备选项。
6. 备选答案的长度要大致一样。
7. 避免使用“以上各项部下”或“以上各项都”之类的字样。
8. 为了对学生的深层思考能力进行测评，可在题干中要求学生对一些隐喻手法进行识别，或者提出一些针对电影剧本的剧情说明，细节、外延涵义的词题。
9. 避免建立备选项的固定模式。
10. 仔细通读所有题目。
11. 将每道题的题干和备选项置于同一页上。

图 8-3 选择题出题指导原则

搭配题可以看成是一系列的选择题，只不过搭配题的题干与备选项并排列出而已。搭配题通常将一组题干置于左列，而稍微偏长的备选项置于右列。学生需要从右边的备选项中选出最相关或者最合适的一项与左列中的每个题干相搭配。搭配题能有效地用于检测学生判定单词、术语或要素间关系的能力。因此，它们在检测学生的认知和理解能力方面最有效。搭配题能要求学生将术语或单词与定义进行搭配，或将简短的问题与答案进行组合，或将一些标志与相对应的名称配对，或者将描述性短语与其他短语、原因与结果、原理与其适用的范围、图片与名称等等进行搭配。

搭配题的使用范围不及选择题广泛，但比正误判断题或者简答题更有效。搭配题的出题和评分工作很简单。由于学生做搭配题的速度较快，因而一次考试中可以使用大量的搭配题，从而可以对更多的知识点进行检测。搭配题最明显的不足之处在于它所考查的是学生的浅层思考能力。搭配题所注重的是学生的记忆力和回忆再现知识的能力，而不是理解能力，而且搭配题常常强调一些琐碎的知识点。

与其他题型一样，出搭配题最好有计划地进行。总体来讲，一组搭配题应该包括 10—20 个思想内容非常相关的题干。如果题干所涉及的话题明显不同，学生就很容易猜测出对应每一个题干的正确答案，因为有的备选项明显与有的题干

内容无关。题干部分应该比备选答案的长度略长一点，应该标上数字序号，而且应该置于左列。备选项应该略短一点，用字母编上序号，并置于右列。备选项的数量应比题干数略多 25% (大约 2-3 个)，而且应在答题要求中指明每个备选项是否允许多次重复使用。正如选择题一样，搭配题中的每个备选项都应是题干的一个看似合理的答案。整道搭配题的所有题干和备选项都应放在同一页上，以免学生将试卷翻来翻去，比较麻烦。最后，答题指导中应明确指出学生该如何标出答案。一般来说，应该在左列题干的数字序号前留出空白以便学生在此标出他们所选的答案。图 8-4 简略地列出了一道出得较好的和一道出得较差的搭配题。

#### 出得较差的搭配题

在左列的每个题干序号前填上右列中与之最合适的表述的相应字母

- |                 |              |
|-----------------|--------------|
| ——1. 亚伯拉罕·林肯    | a. 宾夕法尼亚州的墓地 |
| ——2. 杰斐逊·戴维斯    | b. 买卖奴隶的地方   |
| ——3. 葛底斯堡       | c. 联邦总统      |
| ——4. 约翰·威尔克斯·布什 | d. 美国总统      |
| ——5. 奴隶拍卖场      | e. 被刺杀的林肯    |

#### 出得较好的搭配题

往左列每个描述的序号前填上右列中最合适的被描述对象的相应字母，每个被描述对象可能不止一次被描述

- |                     |               |
|---------------------|---------------|
| ——1. 联邦总统           | a. 亚伯拉罕·林肯    |
| ——2. 被刺杀的林肯         | b. 约翰·威尔克斯·布什 |
| ——3. 同盟将军           | c. 尤利塞斯·S·格兰特 |
| ——4. 在 Bull Run 被刺者 | d. 罗伯特·E·李    |
| ——5. 联盟将军           | e. 杰斐逊·戴维斯    |
|                     | f. 詹姆斯·麦迪森    |

图 8-4 出得较差的和较好的搭配题例子

正误判断题要求学生对一个陈述判断其正确或错误。由于对这种题的解答只有两个选择中的一个，即要么对，要么错。因此，这类题有时又被称为二选一的选择题。

正误判断题经常在教师自行设计的考试中出现。该类题型易于出题和评分，但不太可靠，适用范围也比较有限。正误判断题比较便于教师对大量的教学内容

进行检测，但同时，它所考查的是学生的死记硬背能力和猜测能力，而无法激励他们的理解能力。大多数正误判断题部局限于认知水平的浅层知识点上，部分原因是深层次的表述极少存在明确的正误之分。

正误判断题被认为是检测学生认知水平(再现知识)能力的最有效的方式，图 8-5 中提供了提高其检测效果的指导原则。

正误判断题只有在用于检测学生对大量的知识内容进行回忆时比较有效。然而，为了让正误判断题的使用范围更为广泛，有人建议让学生改正错误的表述或者对自己的答案进行解释。如果附加上这些要求，那么，这种正误判断题就更像是填空题或主观性试题了，从而就不只是检测学生对知识的回忆再现能力了。

1. 先将每个题部写成正确的表述，然后将其中一些改成错误的。
2. 错误的表述应该错在本质性的信息上，而不要在一些琐碎的小事上做文章。
3. 每个题的长度保持一致(错误的题往往被写得比正确的题更短)。
4. 避免直接引述笔记或讲稿中的内容。因为直接引述的题常常只趋于考查学生回忆知识的能力。
5. 避免使用否定的陈述，因为否定的陈述更难写得清楚明了。
6. 谨慎使用一些特殊字眼，如“总是”，“从不”等等。因为很少有“总是”或“从不”正确的事情，因此，善于考试的学生很容易猜到带有这些字眼的陈述多数情况下是错的。同样地，尽量避免使用模糊的字眼，如“有时候”，“通常”等等。因为大多数事情都只有在有时候或者通常情况下才是对的，因此，机灵的学生就会猜到这样的表述是正确的。
7. 在一个陈述中只论及一个主要观点或问题。一道题中所包含的信息太多只会扰乱学生的大脑。
8. 尽可能将至关重要的信息放在陈述的末端，以便使题干表述更加明了。
9. 确保每一个陈述的正误性明确而真实。
10. 正确的题目数与错误的题目数保持大体相当。
11. 避免答案的模式化，例如，对错交替出现。

图 8-5 正误判断题的出题指导原则

填空题要求学生将一个陈述中所缺的部分补充完整或者用一个短语回答题于中所提出的问题。这类题型要么将一个陈述中的关键单词或短语删掉而代之以空格，要么提出一个简单的问题。它既不像其他客观题那样提供备选项，也不像

主观题那样要求学生自行组织答案。

填空题特别适用于年龄较小，已经学会了写字，却不太会自行组织答案的孩子们。填空题比其他客观试题更易于出题，因为它不要求你去想出一些看似合理的备选答案。正如其他客观试题一样，填空题所要求的答题时间较少，因而采用填空题能对大量的知识内容进行检测。此外，由于没有提供备选答案，因而学生无从猜测作答。但另一方面，这类题是典型的只对一些特定的名称或事实进行再现。而且也比较难于做到评分的客观性。例如，一位学生有可能未能填出所期望的标准答案，因为她没能准确地记住某个特定的术语或人名，但她却对这个她忘记了姓名的人做了一番描述。她这个答案该算对还是错呢？

填空题最适合于评价年幼的学生或者对认知层次的知识点进行检测。图 8—6 提供了出好填空题的指导原则。

### 确定考试内容

前面已经指出，最有效的评价应该在教学过程的前期进行规划。不管使用哪类题型的考试，如果你不仅在进行教学规划中，而且在教学过程中都在考虑如何出题的问题，你的做法将无疑是明智之举。因为在进行教学规划中所出的题，由于刚刚确定了教学目标而能够紧扣大纲内容；而教学过程中所出的题能与教学方法和教学中所强调的重点相吻合。在设计一套纸笔测试试题时，效度和信度是首要的考虑因素。你需要确定该考试是否准确地检测了学生所学知识的广度、深度以及学生的思考层次。因此，你应该尽力确保考试的效度，即尽量最大可能使你所教的知识内容与所检测的内容相吻合。

1. 题目要求明确，是同义词的替换，还是纯粹的拼写问题。
2. 确定只有一个可能的正确答案。
3. 避免直接引用教科书或者课堂笔记中的表述。
4. 一个陈述中尽量不要留多个空格，否则会使该陈述表达不清。
5. 要求学生填写一个单词或者简要而确定的短语，而不是一个很长的句子。
6. 直接性的问题一般来讲要比不完整的陈述更加清楚明了。
7. 确定每个陈述都是正确而行依据的\*
8. 将关键性的单词或短语去掉代之以空格。而不是去掉琐碎而不重要的信息。
9. 使用空格时，将空格紧挨着所给的陈述，而放在陈述之后、以便使陈述更加清

楚。

10. 如果所填答案是一个数字, 应在题干的陈述中加上该数字的单位。

图 8-6 填空题出题指导原则

提高考试效度的一种方法就是使用考试蓝图。考试蓝图清楚地描绘出了教学内容主题、认知层次与考试试题之间的关系。表 8-4 列出了一个考试蓝图的例子。从这个表中可以看到, 教学主题被列于左侧纵列中。表的顶部依次排列着 Bloom 分类学中认知目标的 6 个层次。表的最右侧的纵列数据表示出了每一个知识点的考题总数。而表底部的横排数据则表示出了每个认知层次的考题总数。

表 8-4 有关恐龙知识的单元考试蓝图

主题	回忆	理解	应用	分析	综合	评价	合计
恐龙名称	10					10	
	(40%)					(40%)	
生理特征		3				3	
		(12%)				(12%)	
食肉动物	1		2			3	
	(4%)		(8%)			(12%)	
食草动物	1		2			3	
	(4%)		(8%)			(12%)	
灭绝情况	2	3	1			6	
	(8%)	(12%)	(4%)			(24%)	
合计	14	6	5			25	
	(56%)	(24%)	(20%)			(100%)	

在设计一份考题时, 你首先应确定出你希望考试所包含的题目总数, 并把这个数字写在右下角的空格里。然后决定每个主题应该设计多少个题并将数字标在右侧纵列空格里。你的判断应该以各主题的重要性和你教学过程中的重点为依据。对于每一个主题, 你应确定出每一个认知层次需要提出几个问题, 然后将这个数字填在相应的空格中。你可以多出一些比教学大纲中要求检测的认知层次更低的题(因为认知层次越高的题所要求学生运用的秩序性技能越低), 而不要出一些比大纲要求更高认知层次的题。完成考试蓝图之后, 你就按照蓝图中所确定的题数和题型设计考试。如果你早在做教学计划和教学过程中就已经拟好许多考题, 那

么。你只需从这些考题中挑选出最好的部分就可以了。

在表 8-4 所举的例子中，这位教师希望用包含 25 道题的一次考试来检测其学生对有关恐龙知识的一个单元的 5 个主题知识的掌握情况。该蓝图显示出这部分内容的教学重点在于让学生掌握恐龙的名称，因为有 10 个题，即 40% 的考题与这个话题有关。次重点是造成恐龙灭绝的因素（6 个题），其众 3 个主题即生理特性、食草动物和食肉动物所占比重大致一样（各 3 个题）。从该蓝图还看出，如果只要求学生对各种恐龙的名称进行回忆，也许可采用正误判断题或者填主题。而对其他主题则要求学生更深层次的思考，也许可采用选择题或者搭配题。

除了使用考试蓝图，你还可以采取一些别的措施来提高考试的检测效果。图 8-7 提供了一些这方面的建议。

### 管理考试

出完考题之后，你可以通过下列途径提高评价的准确度：“一是尽可能多地向学生提供有关考试性质的信息；二是按照事先定好的考试程序进行。学生应该事先就清楚地知道考试将在什么时候举行，考试所包含的教学内容范围，考题采用了哪些题型，以及考试的评分原则等。这样做能让学生做好更充分的准备。此外，最好建立一个考试程序，然后就一直按照该程序进行每一次考试。考试程序中应包括如下内容：必带的考试材料，考试中教师是否回答学生的问题，学生完成试卷后应该做些什么等等。建立这样的考试程序能让学生对考试有一个连贯一致而且确定的思想准备，从而促使他们更好地把精力集中到考试上，而不至于每次考试都去期待一些新的考试规矩。图 8-8 提供了一些如何管理班级考试的建议。

1. 使用考试蓝图提高考试的效度。
2. 先写出多于你所需要的考题数，然后从中选出与考试蓝图最适应的作为考题。
3. 考试题目中应该包括一些所有学生都能回答正确的问题。
4. 每次考试中只采用 2-3 种题型。对于年幼的学生只采用一种题型。一次考试中的题型过多会混淆学生的思维。
5. 确保考试中的答题要求表说明了通畅。应该让学生知道如何标出自己的答案，知道评分原则以及每道题所占的分值比重等等。
6. 将每一种题型的所有题目放在一起。例如，将选择题的所有题目与填空题分开放置。

7. 留出适当的空白以便卷面更加清晰。每道题之间的空白应该比题干与备选项之间的空白更宽才对。主观性试题要留出足够多的空白以便学生作答。
8. 为了增强考试的信度, 应该尽量增大题量, 但也不能超出学生所能够完成的现实范围。别忘了完成主观性试题要比完成客观性试题所花的时间多得多。
9. 确保每道题的各个部分都位于同一页上。换句话说, 确保每道题的题干和它的备选项都在同一页上。
10. 检查试题以确保没有对正确答案祈祷提示性作用的地方。
11. 留出空白以便考生填写相关信息 (如姓名、日期、科目以及科目编号等等)。
12. 通读试题, 检查答题要求是否表述清晰, 是否有语法错误、印刷错误、拼写错误以及格式上的不正确等。

图 8-7 对考试出题的总体建议

1. 建立考试程序, 将其介绍给学生, 一直按照该程序进行考试。例如应该让学生知道考试开始后是否还可以提问, 考试结束后需要做些什么等等。
2. 所给时间应该足以让所有学生完成考试的各个部分, 不要让一些学生因为时间关系而无法正常发挥自己的实际水平。
3. 在学生进入教室以前就给他们安排好座位。
4. 分发试卷以前, 确保所有学生都有答题所需的工具和材料。
5. 分发试卷以前对考试程序和规则进行解释和重新提醒, 例如, 答完题以后应该做些什么。
6. 控制好试卷的分发方法, 不要让个别学生得到试卷的时间与别的同学相差太远。比较明智的方法是让学生看到你的信号发出后才开始看题或答题。
7. 试卷如有修改之处, 应该在学生开始答题前指出。
8. 避免在学生开始答题后对考试或其中的试题进行讨论。如果你必须提供对考试的补充信息或指令, 或者回答任何一个学生的个别提问, 你都应该大声地面向所有学生进行解释。
9. 考试期间要自始至终地对考场纪律进行监督。

图 8-8 管理班级考试的指导原则

### 学生的表现考评

有时候学生的学习情况很难用纸笔测试的方式进行评价。当教学目标是要求学生创造一幅作品、模仿一套程序性动作、或者表演体育动作时, 最有效的检测

方式就是对他们的产品、程序或者表演的质量进行观察和评价。例如，一位艺术教师可能对学生的作品进行评价，一位口语教师可能对学生的各种发言进行评价，一位体育教师可能对学生的投球技巧进行评价，一位理科教师可能对学生是否按照正确的程序进行一次试验而进行评价，而一位音乐教师则可能让学生演奏一首曲子，从而对他们的表演进行评价。

纸笔测试要求学生完成一项认知性的脑力劳动任务并用一个答案的方式表示出其脑力劳动的结果。而观察性考评则要求学生表演一套动作或者创造出具体的作品，以便让教师对其进行观察和评价。因此，观察性考评常常被认为比传统的考试形式更加真实可靠。正因如此，许多教育学者认为这种真实可靠的考评应该至少代替大部分纸笔测试。

与其他的评价方式一样，观察性考评最令人担心的一点就是所得信息的准确性。虽然整个教学过程中教师对学生的表现在其作品都在进行观察，但这种不正式的观察常常带有随意性。与此相对照的是教师可以仔细认真地、有计划地、按照定好的标准进行正式的观察和评价、从而有利于对学生的表现在其作品成功进行比较。这样做会更具效度和信度。采用书面的检核表、等级评定、描述性等级评价表对学生表现进行正式的观察，是提高观察性考评的效度和信度的常用方法。

### 检核表

检核表是教师用以进行考评的书面工具，表中列出了作为满意的表现所必须具备的基本特点。教师对学生的表现在其作品都在进行观察后就在表中标示出是否具备某一基本特点。然后根据学生表现在其作品所符合检核表各个基本特点的总的情况对学生的整个表现过程进行判断。例如，一位音乐教师可能让学生演奏 3 个音阶，从而对其能力进行评价。图 8-9 给出了对这次演奏进行考评的检核表实例。

评判标准	B 调	F 调	G 调
1. 音调正确	_____	_____	_____
2. 音质良好	_____	_____	_____
3. 节奏连贯	_____	_____	_____
4. 乐器摆放位置适当	_____	_____	_____
5. 姿势自然	_____	_____	_____
6. 气流控制娴熟	_____	_____	_____

图 8-9 音阶演唱检核表

使用检核表的优点在于它能促使教师和学生将其表现的注意力集中在最重要的方面。虽然整个表现过程是很复杂的。但检核表能让学生明确，要完成整个表现过程必须做到哪些事情。同样地，检核表也为每个学生的表现在哪些方面做得好而在哪些方面还比较欠缺提供了特定的反馈信息。因此，学生及其教师通过检核表就可以准确地知道该学生在哪些方面还有待提高。

检核表的最大缺点在于无法对每个基本特点的质量进行评价。检核表只允许教师记录下来某个基本特点是否具有，而无边记录下该基本要素被完成的程度如何。例如，在图 8-9 中教师记录下了一位学生在演奏每个音阶时乐器的摆放位置是否适当。然而，有的同学可能将乐器摆放得更好一些。而检核表无法对这种差异性进行准确反映，从而无法保证评价的高准确度。使用检核表的另一个难点在于如何对整个表现过程进行判断。要达到好的表现程度，难道检核表中的每一项基本要素都是绝对必须的吗？如果是那样的话，学生就必须在表演中对检核表中的每一项基本要素进行展示，而不是展示整个过程。然而，事实上，好像有的基本要素是作为一次好的表现的充分条件，而并非必要条件。如此说来，一个学生的一次表现究竟必须展示哪些基本要素才算是可被接受的呢？这样的问题是很困难解决的。

### 等级评定

等级评定与检核表很相像，它也列出了作为满意的表现所应具备的基本要素。然而，等级评定能让教师对每个基本要素的完成质量进行判断。等级评定不仅能像检核表那样评价出某个基本要素是否完成，而还能评价出这个基本要素被完成的好坏程度或精确度。由此可以看出，等级评定具有检核表所具有的优点，而且基本上避免了检核表的缺点。

有 3 种等级评定类型被广泛运用于班级评价过程中（见图 8-10）。在数字式等级评定中，一个数字表示着表演的质量好坏。一般来讲，越大的数字，反映出其表演质量越好。图表式等级评定要求观察者（指教师）在用以表示学生表现质量的连续性线段上标出记号进行评价。描述性等级评定也是用一条连续性线段对学生的表演进行评价，在这条线段上标有对学生表演、质量不同水平的描述性语言。

用等级评定对学生表现的某个方面或者其整个表现过程进行评判要比用检核表进行评判容易一些。教师只需将各个基本要素所对应的等级进行概括就可做出判断。比较特别的是，经过概括能产生一个单个的数据，而这个单个的数据正

好是该表演整个过程质量好坏的代表尺度。数字式刻度尤其容易进行概括，既可将各刻度值相加，也可计算出刻度平均值。要对图表式刻度或者描述式刻度进行数字化的概括，只需将每个图表刻度或描述刻度标上数字，然后按照数字式等级评定的概括方法加以概括就可以了。有时候需要对学生的表演情况进行更多的描述，尤其在使用图表式和描述式刻度时。在这种情况下，对于表现的每个基本要素的具体描述应该兼顾考虑到每个单一的刻度和对整个表现过程所标示的刻度。

### 描述性等级评价表

描述性等级评价表通常是以一种模板或者表格的形式出现，其内容既囊括了描述式等级评定中的细节内容，检核表中的基本要素特点，又包括了数字式等级评定中的综合性概述。在大多数描述性等级评价表中，每一纵列代表需要评价的某个方面（例如：语法、写作风格、拼写、标点符号等），而每一横排则对每个基本要素被掌握的不同水平层次的特点进行描述。通常情况下，描述性等级评价表对学生表现的每个基本要素的每个水平层次都用一个数字刻度或者一个分数表示出来。

图 8-11 展示出了对学生写作进行评价的描述性等级评价表。4 个纵列表示出了我们希望对学生写作进行评价的四个方面（语法、拼写、标点符号和大写）。针对每个方面列出了 4 行评语，每一行评语描述了该项的一定水平层次。同时，请注意每一纵列最上方的数字和每一横排左端的数字，前者代表每个方面可能得到的最高分数，后者表示每一横排中各个方面所相应的不同水平层次所得的分数。从这些数字我们可以看出，这次写作测验的总分为 40 分，语法、拼写、标点符号和大写 4 个方面所占分值相等，各占总分的 1/4。但这个描述性等级评价表允许教师对学生的分数在每个水平线内灵活处理，每一横排中适用一个分数区间（例如，从 8 至 10 分不等），而究竟如何确定各个区间内的具体分值，描述性等级评价表中就不需要进一步详细说明了。

如果运用描述性等级评价表，教师对学生作业的评分工作会变得更加容易，而且更具一致性。重要的是，学生自己也可用描述性等级评价表对他们自己或者别人的作业进行评价，进而得到更有意义的反馈信息。描述性等级评价表能促使学生和教师都把注意力集中在写作中被认为最重要的方面，正因如此，描述性等级评价表常常与档案袋结合使用。

当你对学生的表演、作品成果、作业进行评判时，等级评定、检核表和描述

性等级评价表能有效提高你评价的准确度。如果它们能满足以下四个条件,那它们的评价基本上是准确无误的。第一,构成良好表现或者可被接受作品的基本要素一定要明确地告诉学生。第二,每个基本要素要具有考察性,并要用清楚易懂的言语进行表述。第三,要指明所评判的表现或作品背景的来龙去脉。第四,所采用的评分或评判等级的程序应该是事先已经确定好的。

数字式等级评定					
根据学生的表演质量在相应的等级数字上画圈 (1=令人不满意, 2=较差, 3=中等水平, 4=较好, 5=优秀)。					
标准	等级				
1. 音调正确	1	2	3	4	5
2. 音质良好	1	2	3	4	5
3. 节奏连贯	1	2	3	4	5
4. 乐器摆放位置适当	1	2	3	4	5
5. 姿势自然	1	2	3	4	5
6. 气流控制娴熟	1	2	3	4	5

图表式等级评定					
根据学生的表演质量在相应的刻度上划勾					
标准	等级				
1. 音调正确	差	中	良	优	
2. 音质良好	差	中	良	优	
3. 节奏连贯	差	中	良	优	
4. 乐器摆放位置适当	差	中	良	优	
5. 姿势自然	差	中	良	优	
6. 气流控制娴熟	差	中	良	优	

描述式等级评定					
标准	等级				
1. 音调准确	部分音调准确	大部分音调准确	所有音调准确		
2. 音质良好	发音含糊, 无力	发音清晰, 但无力	发音清晰有力		

图 8-10 等级评定

	语法 10 分	拼写 10 分	标点符号 10 分	大写字母 10 分
优异 8 至 10 分	所有句子语义完整; 所有动词与宾语搭配适当; 所有名词使用恰当; 所有的句首没有连接词。	每个单词拼写正确; 每个单词书写工整。	每个句子句尾的标点使用正确; 所有的逗号使用恰当; 所有撇号正确使用于表示复数或单数所有格。	每个句子都以大写字母开始; 所有专有名词大写正确。
良好 5 至 7 分	所有句子语义完整; 大多数动词与宾语搭配适当; 所有名词使用恰当; 少数句子的句首使用了连接词。	大多数单词拼写正确; 每个单词书写工整。	大多数句子句尾的标点使用正确; 所有的逗号使用恰当; 大多数所有格用撇号表示。	大多数句子以大写字母开始; 大多数专有名词大写正确。
中等 2 至 4 分	一些句子语义完整; 一些动词与宾语搭配适当; 一些名词使用恰当; 一些句子的句首使用了连接词。	大多数单词拼写正确; 大多数单词书写工整。	一些句子句尾的标点使用正确; 一些逗号使用恰当; 很少使用撇号表示所有格。	一些句子以大写字母开始; 一些专有名词大写正确。
差 0 至 1 分	很少或者没有句子语义完整; 很少或者没有动词与其宾语搭配适当。或者白卷。	很少单词拼写正确; 很少单词书写工整。或者白卷。	很少句子句尾的标点使用正确; 很少逗号使用正确。或者白卷。	很少句子以大写字母开始; 专有名词没有大写。或者白卷。

图 8-11 评价学生书写的描述性等级标准

### 档案袋评价法

档案袋评价法是对学生成绩进行评价的独特而日益普遍的一种方法, 尤其被那些提倡评价要真实可靠的人广泛使用。虽然对档案袋评价法的效度和有用性尚无定论 (Elliot, 1994), 但无论如何, 档案袋评价法已被视为传统的标准化考试的一种不错的代替品。档案袋是指收集在一起的学生学习成果, 这些学习成果能反映出该学生从早些时候到目前这一段时间所取得的进步情况 (Eby, 1992)。在使用档案袋评价法时, 学生得将自己得许多作业收集起来。其中可能包括学生得各种学习成果, 如考试试卷、随堂测验试卷、各种家庭作业、日志以及论文等。你甚至还可以将学生表演的录音、录像带囊括在内。除此之外, 档案袋还可以包括学生对自己作业的评价以及其他学生、教师对该学生作业的评价。由此可以看出, 档案袋是包含各种评价的集合体, 它综合性地提供了反映某学生学习状况及其进步程度的详细而具体的信息。

档案袋评价法的一个关键特征在于, 档案袋评价法是由学生自己参与收集、汇报、取舍、解释反映其学习的各种材料。学生积极参与到对自己的评价工作中, 他们不仅可以决定将自己的哪些作业纳入他们的档案袋中, 而且还要对自己的学习和进步情况进行评价。档案袋评价法的一个首要目的是促使学生更加积极主动,

并以一种主人翁责任感投入到自己的学业中去。

如果你尽量有意让学生及其家长知道档案袋评价法的评价过程,可能使得这种评价更为有效。最好在学年初就与学生讨论他们的档案袋的收集和利用程序。你应该向他们解释说明叫档案袋,档案袋为什么重要以及档案袋有何用处。你还应该允许学生自己确定将哪些材料纳入自己的档案袋中,以便促使他们对自己的档案袋有一种主人翁的责任感。在学年初的讨论中,你应该清楚地申明,哪些作业或学习成果必须纳入档案袋中,而哪些附加材料则由学生自行决定是否将其纳入自己的档案袋里。

档案袋应该保存在学生易于接近的地方,并规定出固定的时间让学生对他们的档案袋内容进行更换或补充。你应该对档案袋进行经常性检查,以确保学生的档案袋中包含有必要的材料,同时对学生的学 习及进步情况进行评价。当你检查档案袋时,你若能对学生的进步情况作些书面批注并将批注放在档案袋里,这会对学生的学 习很有帮助。你还应该鼓励学生偶尔将自己的档案袋带回家以便然让学生家长对自己孩子的学 习进步情况和档案袋的来龙去脉保持长期性关注。如果能像这样利用档案袋,那么,档案袋无疑会是一种有价值的评价方式、一种教育指导工具和交际工具。

(本文摘自《教学行为指导》第八章)

## 评判分数

唐纳德·R·克里克山克 德博拉·L·贝勒尔 金·K·梅特卡夫

在大多数情况下,我们进行班级评价的最后一个步骤就是给每个学生判定一个分数。分数是对学生的表现、一段时间的学习以及各种学习经历进行总结概括的一种简单方式。因此,如果说评价工作包括总结性和形成性评价的话,那么,判定分数就属于典型的总结性过程。最后,你必须综合考虑有关一位学生学习和进步情况的信息,对其进步的快慢做出判断,然后给出一个你认为能准确而公平地代表这个判断的分数。

给学生判定分数是教学过程中很重要的一个环节,然而,通常也是具有威吓性的环节。许多教师反映他们对于给学生确定和报告分数感到担心,甚于害怕(Arasian, 1991; Bradley, 1993)。这种担心一定程度上源于分数在学生及其家长眼里的重要性,以及学校领导把分数看成是反映学生学习进步情况的首要信息。评分工作对于教师来说也绝非易事,因为它迫使教师在两种角色中左右为难,一种角色是施教者,即教学参与者,一种角色是客观判断和观察者,即教学旁观者(Lortie, 1975)。最后,评分工作之所以对于教师具有威吓性,是因为许多教师自认为还没有做好迎接这个挑战的充分心理准备(Hills, 1981)

尽管评分工作困难重重、令人担忧,但分数的用处却不少(Arasian, 1991)。从管理上讲,根据分数能够确定出哪些班级的学生应该受到表彰。学生的进步如何,以及是否应该升入高一年级或者准予毕业。从信息角度讲,分数让家长、学生、以及学校领导得知学生的学习状况和进步程度如何。对于有的学生来说,分数还可以给他们以驱动力,促使他们更加努力地学习。基于上述种种原因可以看出,作为一名教师,有责任选择并实施一套准确、公正一致而且令人信服的评分制度。

在选择和实施一套评分制度时,必须考虑以下 3 方面因素。第一,你在打分时应该综合考虑学生的哪些作业、表现或者测验?第二,在对学生的表现好坏进行判断时你将遵循什么标准?第三,你将如何把来源于好些作业或者材料的信息综合成一个具体分数?让我们对这几个因素进行逐个分析。

### 评分时你应该考虑哪些因素

你所判定的分数意味着对这个学生在各个方面表现好坏的概括总结。你的目的是用一个分数准确地反映出这个学生对所学知识掌握的深度和广度, 所考查的知识还应当与教师在教学过程中所强调的内容相吻合。

一般来讲, 你所给的分数应该包含尽可能多的信息源。信息源越丰富, 你对于学生学习进步情况的反映就越准确。一般的信息源包括考试、小型测验、家庭作业、课堂表现、档案袋和论文等。如果能综合考虑上述多种信息源, 你就能更好地判断出学生的进步情况或者进步趋势。

有的教师还考虑学生的出勤情况, 或者对于学生行为表现、学习态度、课堂表现、学习主动性的非正式观察, 以及学生的努力程度。然而, 一般来讲, 学生的分数最好只包括直接反映学生成绩的因素(Arasian, 1991; Borich, 1992)。如果评分制度中将别的不够清楚、不很确定的因素, 如学习主动性、学习态度等。掺杂到学习成绩中, 必将使分数丧失一致性和正式性。为了解决这个问题, 有的中小学评分制度规定, 给每个学生判定两个分数: 一个分数反映学习成绩, 另一个分数则对学生的学习主动性以及各种行为表现做出判定。如果只给学生判定一个分数, 那它应该只表示学生的成绩。至于有关其他因素的信息可以采用书面批注或者会议的方式传达给学生或者学生家长。

虽然在评分时所结分数包含的信息量越多越理想, 但教师们必须面对现实, 以他们所能够评阅的作业量为限。当然, 将学生整个学期所完成的各种测验的分数都反映到最后的这一个分数中也是完全可行的。然而, 你可能需要有选择性地对待其他的测验分数, 究竟应该把哪几次作业或者测验的分数纳入考虑范围之内。例如, 如果你每天都给学生布置了家庭作业, 每天都有小测验, 要把每次作业和测验的分数都纳入最后分数的考虑之中, 恐怕也是不合情理的。每一位教师都必须决定, 究竟应该把哪些因素纳入学生期末分数的考虑范围之中。

### 你应该遵循什么评分标准

正如前面所说, 评判分数时要求你对学生的学习进步情况做出判断。从我们讨论过的信息源中所获得的信息本身并不能告诉你一个学生的成绩是好还是差。你得将学生的表现与某个标准进行比较才能得出结论。最常用的三种参照标准是:

(1) 其他学生的表现, (2) 事先确定好的固定的成绩标准, (3) 每个学生以前的成绩。

### 与其他学生进行比较

也许你还能回忆起，我们对考试分数的解释中已经说过，如果将一个已知的分数与其他学生的分数进行比较从而做出判断，这就叫常模参照评价。我们用这同一个术语来表述通过将一个学生的成绩与其他学生成绩进行比较而得出的分数。当我们谈到有的教师是根据分数曲线进行评分时，我们的意思是说这位教师采用的是常模参照评分制度。用这种制度评分时，某一个学生所得到的分数是参照同班其他同学完成考试的情况而确定的。如果这个学生完成考试的情况比其他大多数学生都好，那么，他所得的分数就比较高，如果这个学生在考试中比其他大部分学生都完成得差，他就会得一个低分。

常模参照评分制度有两种典型的形式。一种是固定百分比制度 (fixed percentage system)，采用这种评分制度时，一个班的学生中获得某一个特定分数的学生在全部所占的百分比是按一个评分尺度确定好的。一般来讲，这个评分尺度会让大多数学生得“C”，得到“B”或“D”的学生相对少一些，而极少会有人得“A”或者“F”。例如，在下面所列出的固定百分比尺度中确定，学生所得的分数中有 10% 的 A, 10% 的 F, 20% 的 B, 20% 的 D, 40% 的 C。因此，如果用图 8-12 中的原始分数为例，那么，他们所得到的等级分数为，比尔应该得 A，玛丽和琼得 B，卡伦、凯文、帕姆和保罗得 C，伊丽莎白和卡娅得 D，基普得 F。虽然比尔与基普的实际分数只相差 10 分，而最高分数也只是 100 分中的 85 分，他们所得的等级分数看似不合情理，但事实如此，不容置疑。

等级分数	所占百分比
A	10%
B	20%
C	40%
D	20%
F	10%

由此你可以看出，在这种评分制度中，学生所得的分数只能表示出他与其他学生的相对成绩怎样，并不能反映学生自己本身所得原始分数的高低。如果所有的学生都考的不错，那么稍微考差一点的学生就会得到一个较低的分数。因此，在不同测验中的 A 所包含的意义是不等价的。

而另一种更为普遍的常模参照评分制度则是灵活分配制度。在这种制度中，学生所得的分数也是以其他学生为参照对象而得到的，但获得每个等级分数的学

生所占的百分比并未事先确定。取而代之的是，各个等级分数所占的百分比要能够反映出所有学生考试好坏的整体情况。与固定百分比尺度不一样的是，灵活分配尺度中有可能没有一个学生获得某个等级分数。又以图 8-12 中的原始分数为例，所有学生的成绩大致都在同一水平线上。没有一个学生考的特别好，也没有一个学生考得特别差。因此，你可以适当调整评分尺度以便不让学生得到过高的等级分数 A，或者过低的分数 D、F。你可以给原始分数 81 分至 85 分的学生同一个等级分数 B，而给原始分数 76 至 78 的学生同一个等级分数 C。

学生姓名	分数
比尔	85
玛丽	83
琼	83
卡伦	82
凯文	82
帕姆	81
保罗	78
伊丽莎白	76
卡娅	76
基普	75

虽然常模参照评分制度已被广泛采用，然而，这种评分方式的负面影响非常明显，尤其对成绩较差的学生。常模参照评分方式容易打消学生的学习积极性，增大大学生学习压力。导致学生之间的不良竞争，恶化学生学习习惯。进而影响学习效果 (Crooks, 1988)。此外，按照这种评分制度所评的分数很难进行解释说明，因为这种分数反映不出学生对所学科目的掌握情况 (Arasian, 1991)。基于上述原因，尽量避免采用这种评分制度可谓是明智之举。

参照事先确定好的标准 学生的分数还可以参照一个事先确定好的标准而定。这种评分制度属于效标参照评价。在这种评分方式中，一个学生的分数与其他学生的成绩无关，它所反映的是学生对所学知识的掌握情况。通常情况下，这个分数表示出了所得分数与可能获得的最高总分之间的关系。如果一个学生获得了可能获得的最高分数中的绝大部分，他的分数就可算是高分，反之亦然。教师将每个学生的成绩与事先定好的标准比例分数进行对照，从而判定出学生的等级

分数。被普遍采用的效标参照评分尺度如：

百分比分数	等级分数
92-100	A
84-91	B
76-83	C
65-75	D
00-64	F

让我们再一次以图 8-12 中的原始分数为例来看一看，如果采用这种效标参照制度许分法，我们将给出什么样的等级分数。在这个例子中，没有学生能得到 A，比尔能得 B，大多数学生将得到 C，而基普将得 D。从此你可以看出，如果采用效标参照尺度，任何学生可能得到任何等级分数。

效标参照评分制度的目的在于根据客观的标准对学生的学 习进行检测，而不是以其他学生的成绩为参照。因此，按照这种方式评出的分数更能反映学生的学习成绩 (Arasian, 1991)。这种评分制度的最大优点在于每个特定分数表示的含义较清楚。例如一个分数就代表着一个确定的成绩水平，它不依班里其他学生的能力或成绩好坏而改变其固定含义。虽然这种评分尺度是由主观把握的，但采用效标参照评分制度所评出的分数仍被认为相对来说不容易趋于主观偏见。基于上述种种原因可以判定，一般来讲，效标参照评分制是评判等级分数最合适的方式。

### 参照学生以前的成绩

常模参照和效标参照两种评分制度都是将学生的成绩与一个外部标准进行比较，从而评判出该学生的等级分数。然而，这两种评分制没有考虑到学生在各方面的能力和背景知识的差异性。在这两种评分制中，对一位在某一学科里毫无天赋的学生和一位在这一学科里如鱼得水的学生采用同一标准进行评分，这是不够公正的。出此，有的教师认为，自我参照评分制更能以学生为中心，更为可取。在自我参照评分制中，将学生目前的成绩与他过去的成绩进行比较，从而判定出这个学生应得的等级分数。换句话说，学生所得的分数是以他自己的进步情况为依据的。一位进步很快的学生会获得高分，而一位进步很慢的学生就会获得一个较低的分数。这种评分方法常常被广泛运用于协作性学习的情况 (参见第七章)。

自我参照评分制需要事先确定出对于某一个学生来说，需要取得多大进步才能获得某一个特定的等级分数。正如效标参照尺度一样，所提高的不同总分对应

着不同的等级分数。下面所列的尺度显示了依据考分的提高情况评判等级分数的一个自我参照尺度。

比以前提高的分数	所得的等级分数
满分	A
12 分以上	A
9-11 分	B
4-8 分	C
1-3 分	D
没有提高或者有所下降	F

自我参照评分尺度要特殊考虑到那些成绩较好，分数很高的学生。以避免造成对学习成绩好的学生反倒像是一种惩罚的不公平现象。例如，如果一位学生在前一次满分为 100 分的考试中得了 98 分，那么他在这次考试中最多能提高 2 分。如果按照机械的尺度，他即使得了 100 分，也只能得一个 D。然而，根据我们所列的比较合理的尺度，如果这个学生在眼前的考试中获得了满分，虽然只比上次提高了两分，他仍然能得一个 A。尽管如此，也就是说，虽然我们特殊考虑到了以前得分很高的学生的情况，自我参照评分尺度对于以前成绩好、考分高的学生来说仍然是不利的。要保持或者提高一个接近满分的成绩要比在一个很差成绩基础上取得进步困难得多。虽然自我参照评分制避免了常模参照和效标参照评分制所造成的学生之间的不良竞争和评分的死板性，但是，按照自我参照评分制所评出的分数无法反映出学生对这一学科的掌握情况和能力强弱。同样的一个等级分数 A，有可能意味着这个学生对这门学科的掌握情况接近完美的程度，也可能表示一个在上次考试中成绩极差而在这次考试中表现出他对这门学科掌握得勉勉强强。虽然自我参照评分制具有直观性这一点很值得肯定。但基于上述缺陷，可能还是应该尽量避免使用。表 8-5 对上述 3 种评分制度进行了比较。

### 怎样综合评价信息

为了给学生评判一个分数，而这个分数能够代表对这个学生各种表现的总体评判，必须将学生学习过程中的各种表现形式(如各种考试、测验、家庭作业、论文等等)进行综合，最后转化成一个字母，一个数字或者一个符号予以表示。在确定好要对学生的哪些表现进行评分之后，你还必须决定如何对每一种表现进行评分。最后，你必须综合考虑所选择的各种表现的得分，进而得到一个反映该

学生整体学习情况的单个分数。

表 8-5 比较常模参照、效标参照和自我参照 3 种评分制度

	常模参照评分制	效标参照评分制	自我参照评分制
参照对象	别的学生	事先确定好的标准	学生自己以前的成绩
比较方法	根据评分曲线, 确定出获得每个分数的学生比例	根据成绩标准, 确定出学生必须获得多少原始分数才能得到某一个等级分数	根据分数提高标准, 确定学生的成绩必须提高多少才能得到某一个等级分数
分数所代表的含义	该学生与同班其他学生成绩的相对情况	该学生对所学课程内容的掌握情况	该学生相对于前一次考试的进步情况
是否限制每一个等级分数都要被获得	根据评分曲线限制每个等级分数的获得比例	没有限制	没有限制

当你在对所选择的各种表现的得分进行综合时, 别忘了这些分数是对不同的内容、不同的知识点、不同重要性的东西的掌握情况的反映。所有的分数是不等值的。例如, 图 8-13 提供了五名学生的各种学生表现的成绩。从表中你可以看出, 这位教师记录下了学生在下列学习表现中的信息: 两次考试、两次家庭作业, 一次随堂小测验和一篇论文。如果我们只看分数, 每一项表现似乎都同等重要。比尔在第一次考试中所得的 A 与卡伦在随堂小测验中得到的 A 所包含的份量好像没什么差别。然而事实上, 你可能想让两次考试和论文的分数比家庭作业和随堂小测验所占的份量更重一些。

如果你建立一套评分制度, 在这套制度中所有的分数都用同一种方式进行标记, 例如都用字母, 分数, 勾叉, 或者加减符号, 那么, 你将更容易对这些分数进行综合。然而, 有的教师却习惯用不同的符号标记不同类型的作业或考试。在图 8-13 中家庭作业的分数用勾、加号或者减号标识, 而考试、测验和论文的成绩却用字母表示。在这种情况下, 你必须把每一种不同的符号转化成同一种尺度。你可以将所有这些符号都转化成数字。利用图 8-13 中的信息时, 你可以将家庭

作业这一项中的成绩这样转化成数字：加号转化成 5 分，勾 3 分，减号 1 分。类似地，你可以将字母分数都转化成数字分数；A 转化为 5 分，B4 分，C3 分，D2 分，F1 分。现在，所有的信息都转化成了数字的形式，从而能够进行代数综合，通常只需算出平均分数就行了。然而，在此之前，你必须确定其中每个分数在最后的总体性得分中的重要性。例如，论文可能是最重要的，考试次之，随后是家庭作业和随堂小测验。

大多数教师都会在对学生的各种分数综合之前对其进行增值处理。为此，他们将每个单项成绩分数按照其重要性乘以一个相应的数字。在利用图 8-13 的数字信息时、每个单个的考试分数可以乘以 2，每个论文成绩则可以乘以 3。如此增值处理之后所得的分数用数字表示如图 8-14 所示。现在你就可以将这些经过增值处理或调整后的分数用于进行最后综合处理了。

学生姓名	家庭作业 1	家庭作业 2	随堂测验	考试 1	考试 2	论文
比尔	✓ (3)	+ (5)	B (4)	A (5)	A (5)	A (5)
琼	✓ (3)	✓ (3)	B (4)	C (3)	B (4)	A (5)
卡伦	+ (5)	✓ (3)	A (5)	B (4)	B (4)	B (4)
伊丽莎白	+ (5)	✓ (3)	A (5)	B (4)	B (4)	A (5)
卡娅	- (1)	✓ (3)	D (2)	C (3)	B (4)	C (3)

图 8-13 学期分数抽样表

当所有的分数都已经被转化成同一种尺度，通常转化成数字分数，而且这些分数已经经过增值处理，能够反映出各种成绩的不同重要性以后，你就可以将它们综合成一个单个的分数了。

为此，你得将增值处理后的所有单项成绩相加得到一个总分数。然后用这个总分数除以被增值处理后的成绩项目数。在我们所举的例子中，最开始的成绩项目数为 6，但是我们对有的项目进行了增值处理。增值处理后的成绩项目数就变成了 10，因为我们将两次考试中的成数按 2 倍计算，而论文的成绩则按 3 倍计算。

增值处理后的成绩项目数算法如下：

家庭作业 1 =1

家庭作业 2 =1

随堂小测验 =1

考试 1 (×2)	=2
考试 2 (×2)	=2
论文 (×3)	=3
增值处理后的成绩项目总数	=10

用学生所得的各项增值处理后的分数总和除以增值处理后的成绩项目数, 就得到该学生的最后平均分。将学生的最后平均分与我们开始用来将各种符号和字母转换成数字的 5 分制尺度相比较就可以给该学生评判等级分数了。图 8-14 中列出了 5 位学生各自的最后平均分和等级分数。

学生姓名	家庭作业 1	家庭作业 2	随堂测验	考试 1	考试 2	论文	平均分数
比尔	3	5	4	10	10	15	4.7 (A)
琼	3	3	4	6	8	15	3.9 (B)
卡伦	5	3	5	8	8	12	4.1 (B)
伊丽莎白	5	3	5	8	8	15	4.4 (B)
卡娅	1	3	2	6	8	9	2.9 (C)

图 8-14 增值处理后转化为数字分数的学期分数抽样

评分定级工作是教学过程中令人望而生畏的一个环节, 但同时也是十分重要的一个环节。它通常是你整个评价过程的最后结果。如果计划得当、实施准确, 评分定级工作就会是一个公正、准确而且有用的过程, 它能为学生的学习和表现情况提供有效的信息。

### 最后的几点思考

对学生的学习情况进行评价是有效的教学过程中至关重要的环节之一。这样做不仅有利于教师把握住学生对所学知识掌握的多少好坏, 而且有助于教师明确自身教学工作的效果怎样。要对学生的学习进行准确的评价需要教师从各种各样的信息源中收集有效、可靠的信息, 然后根据这些信息做出判断。评估信息最被公认的源泉就是考试, 包括标准化考试和教师自行设计的考试。但大多数教师还会采用大量别的信息源, 包括正式的和非正式的。他们最终的目的是, 对学生的学习成绩做出一个公正、一致、有效的判断, 以便向学生及其家长、教师自己以及别的教育专业人员提供有价值的反馈信息。

(本文摘自《教学行为指导》第八章)

# 基于多元智能理论的大学生学业评价模式构建

董 勇

浙江工业大学 之江学院, 浙江 杭州 310024

**摘要:** 本文从多元智能理论视角论述了对大学生学业评价的价值导向, 以及基于多智能理论构建大学生学业评价模式的原则, 通过对评价准备、评价实施、评价反馈三个阶段的内容与方法的分析, 探讨了构建多元评价主体参与的, 旨在促进学生发展的大学生学业评价模式。

**关键词:** 多元智能理论; 大学生; 学业评价模式; 构建

学业评价是教育活动的重要内容。促进大学生认知、能力、情感全面发展, 是开展学业评价的出发点与终极目标, 因此, 学业评价须改变过去过份强调的甄别与选拔功能, 充分发挥促进学生发展、教师提高和改进教学实践的功能。在这样的评价理念指导下, 大学生学业评价不再只是传统意义上的检验环节, 而是教育活动的有机组成部分, 贯穿始终。通过评价与反思、反馈与修正, 促进学生知识阐释、批判性思维和自我完善。要实现这样的目标, 构建新型的大学生学业评价模式是关键。而多元智能理论是构建大学生学业评价模式的理论基础。

## 一、多元智能理论对大学生学业评价的价值导向

### (一) 从一元到多元

1983年, 美国哈佛大学教授霍华德·加德纳在《智能的结构》(Frames of Mind)一书中首次提出多元智能理论, 认为人类有着完整的智能“光谱”。这一论断突破了传统智力理论的假设: 人类的认知是一元的, 可采用单一的、量化的智力检测手段来测量人的智能。经过多年的研究, 加德纳逐渐完善了这一理论, 明确提出人类存在多种不同的思维方式, 他将人类的智能类型分成9种, 分别是: 语言智能、逻辑数学智能、空间智能、身体运动智能、音乐智能、人际关系智能、自我认识智能、自然观察者智能及存在智能。多元的智能观是对一元的突破, 是片面向全面的回归, 它改变了传统的智力=语言智能+数学逻辑智能的思维定式, 使我们对人类的智能有了更深层次的认识。每个人的智能是多元的, 虽然并非每一个方面都能达到最高水平, 但有一方面是最佳的, 并且每个人的最佳智能单元是平等的。实际上, 人的智力是一种潜能, 每个人都具备多元的潜能, 都具有可挖掘的创造潜

力。这促使我们走出以片面的、单一的标准评价学生的误区,开始以多元的、全面的视角去分析、理解和引导学生,以宽容的心态接纳和信任每个学生。

### (二) 从知识到能力

长期以来,在传统智力观的影响下,在教育活动中占主导地位的是认知能力的培养,强调知识的学习与积累。而多元智能理论认为,教育的目的在于使学生得到最大程度的理解。什么是“理解”?加德纳在1991年出版的《未受教育的心理》一书中指出,他或她若能将在任何教育背景下所获得的知识、概念和技能,应用到与这些知识确实相关的新的事件中或新的领域内,那么他或她就具有真正的理解并学以致用的能力。在他看来,智能是在特定的文化背景下或社会中,解决问题或制造产品的能力,即能够针对某一特定的目标,找到通向这一目标的特定路线。多元智能理论的教育观更多地是注重能力的培养,从注重知识的传授转向注重能力的提升,是多元智能理论的重要价值取向。这启示我们,学习是学习者建构知识结构或者心理表征的主动加工过程。学生不但对教师所教的学科内容建构起这样的结构,他们还会建构起用于思考这些信息的策略、在意识层面感知这些策略的方法以及促进策略使用的途径。

### (三) 从统一到差异

自从心理学家阿尔弗莱德·比奈(Alfred Binet)发明“智力测验”以来,智商测试或学习能力倾向测验(SAT)一度成为备受推崇的智力测量工具。在工具理性盛行的时代,人们期望通过精密复杂的智商测试来判断一个人的智力高低,根据测试成绩对学生进行令人可信的排序,最聪明的和最有前途的被接受最好的教育。建立在这种观点上的统一规划的学校里,长期以来占主导地位的教育采用的是“工厂模式”,每个学生学习相同的课程,被按照统一格式培养,接受同样的纸笔测验的评价。在多元智能理论看来,因为人类存在许多不同的、相互独立的认知能力和认识方式,所以,在每一个人的身上,都是数种智能的不同组合。每个学生都有着自己的优势智力领域,有自己的学习类型和方法,学校不存在“差生”,每个学生都是各有智力特点、学习类型和发展方向的可造就人才。从这个意义出发,学校应该强调学生的差异性,学校的教育活动应冲出统一规划的樊篱,充分体现学生的个体差异。

### (四) 从规范到引导

多元智能理论认为,每个人都有自己的智能强项和智能弱项,表现为与之相

应的独特的认知方式。教师要根据每个学生的认知特点进行合理有效的教学,使他们的优势智能得到充分的展示,再将优势智能领域的特点迁移到弱势智能领域中去。教师在教学中的角色是多重的,既是一个观察者,又是一个研究者,更是一个引导者。在教与学的过程中,教师应注重观察,通过观察学生的学习认知加工过程和思维过程,发现学生在教学过程中展示出的智力特点,发现学生的潜能所在。在此基础上,针对不同的教育对象,选择恰当的教学内容与教学策略,通过建构合适教学环境,帮助学生认识自我、发展自我,着重激发学生的潜能,促进学生身心的全面发展。在多元智能理论的视野中,对于学生个体差异的敏感程度,应该成为教师能力的一部分,并运用于正常的教学中。教师不再是教学规范的执行者,而是教学过程的引导者;学生不再是被动的知识接受者,而是学习的建构者。

#### (五) 从选拔甄别到重在激励

传统的纸笔测验,是特别注重量化指标、信奉教育效率的结果。由于考试的认证、甄别与选拔功能可以为个体进入社会生活提供相应的学业成就水平的证明,所以在我们的日常生活中除了学生的标准化考试之外,还有用于评价教师、督学、士兵和公务员的各类测试。考试不仅用于便于测试的内容,如物理、数学,也同样用于不太便于测试的内容,如艺术、文学,加德纳称之为“完全考试化的社会”。可是,面对学生个体之间在观念、兴趣、爱好、思维方式和现有知识结构等方面呈现出的差异性,多元智能理论认为,在学生潜能开发、创新精神和实践能力培养的过程中要尽可能运用灵活多样的教育手段,如运用多样化的情景,采用多样化的活动方式特别是应用多样化的评价手段等来正确引导和促进不同个体潜能的开发。评价的功能更应注重激励与导向,通过评价,关注学生的学习变化及成长历程;通过评价,激发学生的潜能,促进学生智能得到整体的、充分的发展。

## 二、基于多元智能理论的大学生学业评价模式构建

20多年前美国教育促进委员会(National Commission on Excellence in Education)针对美国学校发表了一份题为《国家处于危险之中》(A Nation at Risk)的报告。这份报告列举了美国学校的几大失败之处,其中之一是没有能够教会学生如何进行批判性思维及如何应用所学的知识。随后,相关的研究报告相继涌现,并引发了教学改革的热潮,这些改革运动导向学业评价的新观念更新。

学业评价指在一定教育价值观指导下,根据一定的标准,运用现代教育评价的一系列方法和技术,对学生的思想品德、学业成绩、身心素质、情感态度等的

发展过程和状况进行价值判断的活动,其根本目的是促进学生的全面发展。

在当前的大学教学实践活动中,学生学业评价根据课程性质的差异略有不同,理论教学课通常采用考试这样的终结性评价方式,通过“教学-期中考试-教学-期末考试”的流程开展教学,通过考试分数来进行学生评价,更多的是注重检验、诊断教学成效;实践教学课通常采用结果评分的方式,通过查看学生的实验结果,由教师给出相应的评价,通过量化的评价标准来判断学生的学习是否达标。不难看出,无论是理论课还是实践课,评价主体是教师,学生只是作为评价客体被动地参与评价过程;评价的内容只涉及学习,最后是通过分数的高低、排名的先后来评价学生,评价关注的重点是教学效果,即学生的学习是否达到预期的教学目的。这样的评价方式不可避免地带有应试教育的影子,按照美国学者洛克希德(Lockheed, M. E.)的观点,传统的学生学业评价有六个最普遍的目标:为高一级的教育选拔学生;认证学生的成就;监测成就变化的趋向;评价特定的教育项目和政策;促使学校、学区对学生成就负责;诊断个体的学习需要。这些指标基本上都属于“对学习的评价”。

当前教育评价领域更多地关注“为学习的评价”。布卢姆的教育目标分类学表明教育者开始清楚地表达对这种评价的需求——专门设计教育目标并将该目标用于计划、教学、学习和评价这一循环圈中。强调学业评价应构建一种支持学习的评价模型。因此,在多元智能理论的指导下,对大学生学业评价模式构建进行探讨,将有助于实施有效评价,有助于实现支持学习、促进学生发展的评价目标。

### (一) 大学生学业评价模式的构建原则

1. 开放性原则。评价主体多元、开放,因而评价不再是教师的专利。由教师、学生、家庭、社区构成多维度的评价主体,可以从多个方面、多个角度出发对教育活动进行更全面、更客观、更科学的评价。评价主体由单维度转向多维度,使得学生由原先的评价对象成为评价主体,学生不再处于过去单纯的被动状态而是处于一种主动的积极参与状态,通过自我评价充分体现他们在教育评价活动中的主体地位,强化学生的自我认识,这有利于教师学生不断地对自己的教与学活动进行反思,进行自我调控、自我完善、自我修正,进而不断提高教育活动的质量和效率。

2. 过程性原则。在评价过程中,须突破传统学业评价只注重结果和只在学习结束时进行评价的框架,注重考虑被评价者的变化发展动态及其过程。更注重在

学生进行学习或研究的过程之中进行评价,重视学生在自主学习或自主研究过程中表现出来的情感、态度和内心体验,要重视学生在发现问题、分析和解决问题过程中的知识能力与见解创新,着重考察学生“是否做了”及“怎么做的”。

3. 发展性原则。发展性原则是指学业评价应着眼于学生的进步与发展,着眼于教师的教学改进和能力提高。多元智能理论主张“评估的观点”要看重创造与思考技能的开发,强调“要通过情境化的评估激发智能”。现代教学论认为,任何一门学科的教学目标都应指向促进学生的综合素质提高以及个性与特长的发展,都含有基础性学力和发展性学力两个层面。严格意义上讲,教学及其评价的目的都在于提升人的生命质量,促进人的全面发展。

## (二) 大学生学业评价模式

构建基于多元智能理论的大学生学业评价模式,旨在通过多维主体在评价阶段的全程参与,通过评价准备、评价实施、评价反馈三个阶段的工作,发挥评价主体各自不同的作用,达到激发学生潜能、促进学生发展的目的。基于多元智能理论构建的大学生学业评价不再局限于单向的、机械的考试这一方式和手段,而是一个多元主体的、循环往复的、不断促进学生提高的系统(见图1)。

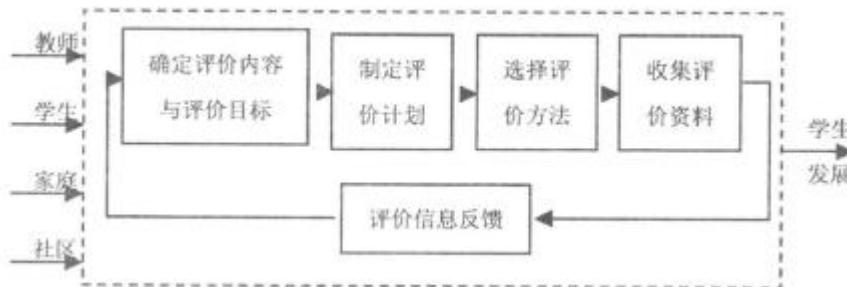


图1 基于多元智能理论的大学生学业评价模式

在该评价模式中,大学生学业评价相关各方:教师、学生、家长、社区等均是评价主体,共同参与大学生学业评价,通过多元主体的交互作用,构建适宜的人际环境和心理环境。该评价模式不仅关注评价结果,更关注评价资料的收集、反思与反馈;不仅关注评价的公正、全面与客观,更关注学生的发展过程和发展潜能。在该评价模式中,评价主体间是战略伙伴关系,学生在参与对自己的评价过程中,成为对自己发展负责的人;学生之间通过相互评价,促进小组合作学习,使每个学生看到他人的优势,反思自己的不足;教师在评价中不仅是知识的传授者,还是引导者与陪伴者,在共同评价的过程中教师可以发现学生的智力差异,从而

修正自己的教学策略；家庭和社区也不再是学生发展的旁观者，而是积极参与其中，为教学和大学生学业评价提供重要的支持。

### 三、基于多元智能理论的大学生学业评价模式过程实施

#### (一) 评价准备阶段

1. 确定评价内容与标准。传统的学业评价更多地注重语言智能与数学逻辑智能，而基于多元智能理论的大学生学业评价内容涵盖更广，既要关注知识的掌握、智力的发展等认知领域，关注学生思想品德、个性、人格等情感领域，也要关注创新精神与实践能力素质。可以采用的评价指标体系有：“素质要素模式”，从“德、智、体、美”或道德品质、公民素养、学习能力、交流与合作、运动与健康、审美与表现等六项“基础性素质”来展开；“发展要求模式”，用若干个“学会”来表述；“行为要求模式”，用“勤奋学习”、“遵纪守法”等行为化指标来表述；“重点特色模式”，由联合国教科文组织提出的“学会认知”、“学会做事”、“学会共同生活”、“学会生存”等“四个支柱”构成。

2. 制定评价计划。评价计划的制定环节，由评价主体共同协商、研究全面发展性评价目标，并将评价目标分解达到可操作化程度。评价主体共同研究评价目标、周期、实施、活动、措施、检查、反馈等，对不同情况、不同类型、不同程度的学生，制定不同的评价计划。

#### (二) 评价实施阶段

1. 选择评价方法。终结性评价与形成性评价是常用的评价方法。两者之间的区别在于，终结性评价注重量化评价，通过对评价指标体系的量化，获得评价数据，评价结果的数据化使得评价结果更容易被统计、测量与排序；形成性评价注重质性评价，对评价对象的那些能够显现事物内在规律的具体特征进行细致观察和深度挖掘，以归纳模式对观察、访谈等形式收集到的信息进行分类与凝炼，并进行解释和价值判断的一种评价。无论那种评价都是为目的服务的，选择何种方法进行大学生学业评价，应根据实际情况进行。但评价方法选择要符合学生的身心发展规律，概括而言，学生的身心发展规律有以下四个特征：①发展是一种自我建构；②个人的发展是在与他人的交往、与环境的互动中实现的；③发展是一个长期渐进的过程（不排除某些飞跃）；④发展是有差异的，既有各个发展阶段的不同特点，又有各个个体的独有特征。这些特征启示我们，在大学生学业评价实施过程中，评价必须调动、依靠学生自身的主体积极性，评价方法的选择在不同的情境下，

在学生发展的不同阶段,可以灵活地对评价方法进行组合。只有这样,才能有效地描述学生全面发展的状况,也才能评定复杂的教育现象。

2. 收集评价资料。收集评价资料是全面评价学生的基础。这些资料通常包括学生的自我评价、教师和同伴的观察与评价、来自家长的信息、来自社会的反馈、考试和测验的信息、成绩与作品集、其他有关说明学生进步的材料等。常用的收集方法有:(1)成长记录档案。在整个学习过程中的有目的地收集学生特定领域的努力、进步和成就的相关资料。如:①展示型成长记录档案,主要收集能够反映个人所取得的成就材料;②过程型成长记录档案,主要收集反映不同时间、阶段的个人表现材料。无论哪种类型的成长记录档案,其中都应该包含本人对所收集材料的反思。如为什么选择这些材料的说明,对每份材料所蕴含意义的解释,伴随每份材料的学习经历、情感体验、学习方法、学习体会等。(2)观察与访谈。观察法是在不进行人为干预的前提下,通过观察将学生在教学活动中实际呈现的现象及其过程客观地记录下来,适用于难以量化评价的内容;访谈法则是通过与当事人和与其有关的人员(如家长、教师、同学等)的访谈从侧面获取学生的信息。观察与访谈有助于全面了解学生在学校、家庭、社区的真实表现。(3)常规测验法。测验法是传统学业评价中运用最广泛的方法,也是现代教学评价中较常使用的方法。测验法适用于便于量化的评价内容,它可较大程度地反映出被评价者的知识掌握情况、能力水平等。测验可通过笔试、口试、论文等形式进行。为了保障测验结果的有效性,要把握好测验的质量指标:信度、效度、难度和区分度。(4)小组合作评价。将社会心理学的合作原理纳入教学之中,强调人际交往对于认知发展的促进功能。小组合作评价的特征是“两个结合,两个侧重”,即学习过程评价与学习结果评价相结合,侧重于对过程的评价;对合作小组集体的评价与对小组成员个人的评价相结合,侧重于对小组集体的评价。(5)真实性评价。真实性评价是指评价不仅要考查学生获得了哪些知识和技能,而且要考查学生对知识和技能的应用能力。通过提供给学生与现实生活相关领域类似的真实性任务,让每个学生充分应用相关知识、技能以及策略,表现其理解水平和对知识的应用能力。真实性评价通过在真实情境中对学生的考查,提供被评价学生的真实信息。

我们应该认识到,资料收集的过程就是评价的过程,是学生、教师不断自我完善的过程。通过各种方法收集的资料不仅应涵盖学生发展的优势领域,也应涵盖学生发展不足的领域,这样才能为学生的发展建立全面的、客观的资料档案,清晰

描绘出在教学活动中学生成长、发展的轨迹。在收集材料的过程中,评价主客体之间通过自觉的、自愿的、民主的、和谐的促进主客体共同发展的行为过程来实施评价,通过发挥多主体力量,全员民主参与、全体多边互动来实现评价目标。

### (三) 评价反馈阶段

1. 评价材料分析。评价主客体共同针对评价过程阶段收集的材料进行全面客观的分析,把绝对标准(统一的预先制定的目标)、相对标准(在学生内部确定一个基数进行横向比较)和个体内差标准(将学生的过去表现、常态表现和理想表现进行比较)结合起来,以个体内差标准为主来进行综合分析。通过对学生学业成绩、道德品质、心理素质、人际交往能力、学习兴趣、积极情感体验等方面发展的评价与反思,发现学生在发展中的优势领域,寻找学生在发展中的劣势领域;总结学生在评价过程中做出的成绩、值得推广的经验,发现学生在评价过程中表现出的问题;找出导致问题出现的原因并提出改进方向。

在评价分析阶段,评价主客体之间可以对评价实施过程中收集的材料,结合评价标准积极开展自评、互评与他评,通过客观真实的、主体多维的、内容丰富的、标准多元的评价发现学生特长,展示学生才华,激发学生潜能。通过评价分析帮助学生正确地认识自己、悦纳自己、拥有自信。

2. 评价信息反馈。评价信息的反馈是下一轮评价实施的开始。针对学生发展中存在的问题,评价主客体共同分析、讨论、协商,修正、增补、调整评价计划,确立新的评价标准,为进入下一阶段的学业评价作好准备,并开始收集再评价材料。通过评价反馈使学业评价模式成为一个良好的闭环系统,在评价-反馈-评价的过程中,鼓励学生不断地完善自己,扬长补短,促进自身的全面发展。

学生作为评价主体之一,在整个评价实施过程中都是积极的参与者。和谐、友好、民主的评价氛围,可以帮助学生更好地接纳认同评价结果,有助于教师对学生发展的激励与引导。利用评价信息反馈可以促使学生自我构建新的目标体系,针对新目标,评价主客体共同研究制定新方案、新计划、新措施,并开始新的评价程序。通过共同合作,评价主客体履行各自的评价任务和职责,使评价各方进入一个相互学习、相互帮助、相互促进、相互发展的新阶段。

(本文摘自《浙江工业大学学报》2009年第1期)

传教 学工道

引教 学工路

浙江工业大学教师教学发展中心

地址：浙江省杭州市潮王路 18 号存中东楼

邮编：310014

电话：0571-88871056

网址：[www.jxzx.zjut.edu.cn](http://www.jxzx.zjut.edu.cn)

邮箱：[jiaoxzx@zjut.edu.cn](mailto:jiaoxzx@zjut.edu.cn)

教学爱好者 QQ 群：150921474



教师教学发展中心  
Center for Faculty Development