●国家基础教育课程改革系列丛书

世界课程改革与教学创新

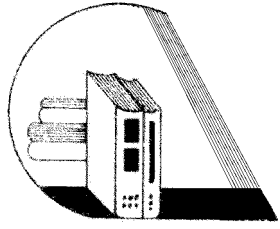
文　　库

（第二辑）

课堂教学改革的理论与方法

各国(地区)学业评价缺席与方法

北京师联教育科学研究所　　编



學苑音像出版社

需要全本请访问 www.ertongbook.

责任编辑:王　军
封面设计:师联平面工作室

世界课程改革与教学创新文库
(第二辑)
课堂教学改革的理论与方法
各国(地区)学业评价制度与方法

北京师联教育科学研究所　编
学苑音像出版社出版发行
★
三河文阁印刷厂印刷
2004 年 8 月第 1 版　第 1 次印刷
开本:850×1168　1/32　印张:157　字数:4060 千字

ISBN 7-88050-122-3

本书配碟发行全 40 册 640.00 元(册均 16.00 元 不含碟)

本书如有印刷、装订错误,请与本社联系调换

目　　录

国内外学业成绩评价制度 ……（1）
国外的教学检查和评价 ……（7）
国外学校考试与评价理论的发展 ……（14）
当前国外学校评分制度 ……（18）
国外学生成绩评定的一些特点 ……（25）
以学生的学习方法为基础的教学评价 ……（30）
国外对评分制的批判与改革 ……（34）
美、德、法、苏中小学成绩评定的若干特点及其问题……（43）
美国中小学评分方法的类型与特点 ……（52）
美国的标准化测试 ……（56）
美国的‘中位数’评定方法 ……（61）
美国少数民族学生测验与评定的产生、发展与种族主义教育 …（65）
美国多元文化教育中测验的目的及其问题 ……（69）
英国中小学综合评估模式介绍及分析 ……（74）
课程作业——英国中小学生学业评定的主要形式 ……（78）
英国国家课程和学业评定 ……（82）
英国学者对高等教育教学质量评估可信度的讨论 ……（96）
东欧国家评定学业成绩的基本概念和趋势……（102）
几个东欧国家中小学学业评定办法……（104）
澳大利亚国立大学考试方法……（111）
“成绩报告单”与日本小学的评价制度……（116）
日本中小学教育评价……（124）

考试形式、类型与方法的发展 …… (126)
考试制度的改革与完善 …… (129)
当前考试改革的趋势及其展望 …… (132)
夸美纽斯对西方教育考试制度发展的贡献 …… (134)
罗伯特·蒙哥马利考试的新探索 …… (139)
考试的历史背景 …… (142)
考试的功用 …… (152)
科学考试法 …… (166)
中外考试的过去、现在和将来 …… (182)
考试的可比性 …… (192)
关于考试影响问题的思考 …… (199)
在考试中引入行为评价 …… (205)
让学生带一张卡片参加考试 …… (212)
如何衡量考试效果 …… (214)
标准考试与真正的学习 …… (216)
考试的改革 …… (220)
中小学考试升学制度的构成、生成与变革 …… (222)

国内外学业成绩评价制度

不同的国家，不同年代乃至不同的学校对学生的学业成绩评定可能会采取不同的方法。我国自汉代开始，学校教育中普遍采用等级记分法。如汉代太学的成绩评定分成及格和不及格两个等级，后又分甲、乙、丙、丁四等或优、良、中、可、劣五等。中国学校采用百分制最早始于清代末年，1902 年（清光绪二十八年）清政府颁布的《钦定学堂章程》中规定："评定分数以百分为满格，通过各科平均计算，每科得 60 分者为及格，不及 60 分者为不及格。"百分制记分法断断续续地一直沿用至今。新中国成立后，起初学习苏联广用五分制，后改百分制。"文革"期间，学校曾用及格与不及格两级制以及百分制等方法。现在我国学校教育习惯上采用百分制记分方法，60 分及格似乎天经地义。尽管有时人们也采用"及格与不及格"两级记分法以及"优、良、中、及格、不及格"五级记分法，但百分制毕竟是人们最常使用，同时也是最熟悉的一种记分法。

评定学生成绩的方法多种多样，太原师范学院孙晓雯、郭海荣老师研究认为，从外在表现形式上可分为字母和数字评价制度与描述性评价制度。

一、字母和数字评价制度

等级分制中包括字母等级分和数字等级分，可以用字母表示也可以用数字表示。如五分制，可用 A、B、C、D、F(E)表示，也可用 1、2、3、4、5 表示。采用几级分制，各国的情况不同。

需要全本请访问 www.ertongbook.

1. 字母等级分制

在国外，字母等级分是最常用的一种记分形式。美国的高中75%采用的是字母等级分 A－F 五级分制，通过在字母后加＋或－号，如 A＋或 A－，这种评分形式可以有 75 个等级。由于使用的广泛，它被社会所接受。

这种记分形式的缺陷为：(1)标准不统一。学校与学校、学区与学区间的分数质量不同，一个学校的 A 也可能只相当于另一个学校的 B 或 C。(2)它不能清楚地反映学生的实际学习状况。(3)无法计算学生的平均成绩。

2. 数字等级分制

这种记分形式实际是百分制的变形，在计算时采用 100 分制，在给学生成绩时采用字母等级分表示。A(90－100)，B(80－89)，C(70－79)，D(60－69)，F(60 以下)。也可以用数字 1～5 来表示分数等级，或者采用 10 分、20 分等级。

3. 及格——不及格等级分制

目前采用这种记分方式的学校已经比较少了。因为这种记分制所含的信息量太少，学生无法了解自己的实际程度，不知从何改进，也不利于教师和学校的管理工作。在两级分制中实际很少有学生得不及格，它的评价意义就失去了。

4. 数字百分制

从教育测量运动开始，定量评价被广泛应用，在美国仍有一些学校采用百分制，但数量很少。在校内学生成绩评价中百分制已基本不被采用，原因是百分制在对学生进行评价时存在很多缺点：其一是当测验的目标不止一个时，相同分数不能直接进行比较。其二，在难度不同的各测试之间不能进行比较。其三，看似精确的分数实际并不能反映出学生真实的差异，由于试题本身的偶然性，学生在考试中 10 分的差距没有实质性的不同。因此，各国大都不采用百分制。

二、描述性评分制度

1. 考核表

通常考核表作为字母等级分和数字等级分的补充,由于前两种记分形式都无法反映学生的一些非智力的因素,采用考核表可以较全面地评价学生。它把学生成绩的若干方面如知识、行为、态度、动机、能力等制定成指标,并把这些指标分成等级,教师由此评定学生是否达到各项指标,其等级如何。

考核表不足之处为,需要制定详尽的评价指标,增加了教师的工作量,不易操作。

2. 评语

评语是目前被广泛使用的评价方法。通常在学生的成绩册上,教师要附加上评语。目前,对于学生的动机、兴趣、努力等非智力因素尚无公认的可操作的指标,对此教师多用评语的形式给学生以评价。一般评语不单独使用,而是配合学科成绩分数给学生以全面的评价。

3. 学习表现评价

学习表现评价是 1989 年由 Resnic 提出的,这种评价学生的方法要求采用口头报告、扩展项目、工作量、科学调查、创造作品和体育奖牌等方面来评价学生。通过这些具有明显标志的作业,来确定所要求的知识是否已经在学生的学习过程中成功地转化成了“综合表现”。

学习表现评价强调学习的认知过程,包括决策、问题解决、独立判断和创造性,因此它与其他评价方法有很多共同之处,它的特点在于它对学生的评价是采用评语形式而不是分数。

4. 学生作业抽样

这是一种用学生作业代替分数的做法。学生把他们作业中的一份带回去给家长看,使家长能够了解学生的学习情况。这份作业就

是学生的成绩。这种评价制度倾向于强调教育必须适应能使学生获得成果的学习需要。它的优点为家长可以根据学生的学习结果(作业)评价学生的学习状况,教师减轻了用分数评价学生的责任。这种评分方法的缺陷为,很难选择适当的作业,另外家长不易评分。

从教育测量与评价的专业角度来看,成绩评定等级制可分两大类,一是绝对评价等级制,二是相对评价等级制。

按照既定的学业标准把学生的学业成绩与表现划分为若干等级。这种成绩评定法称为绝对评价等级制。常见的有如下几种:

1. 教师综合评价法

这种方法是由教师对每位学生的情况做全面考察,包括平时作业,课堂提问与讨论、形成性测验结果、学生提供的作品等多方面的学习表现与成果,对学生的学业成绩进行综合评价并给出等级分数。

2. 测量结果转换法

这种方法通常是在举行正式测验或考试之后,把测量得到的分数按照事先规定的分数段标准转换成等级分数。例如,90 ~ 100 分为优秀(5 分),80 ~ 89 分为良好(4 分),70 ~ 79 分为中等(3 分),60 ~ 69 分为及格(2 分)。59 分以下为不及格(1 分)。有些国家使用五级记分制,有些国家则采用七级记分制。

3. 核心内容参照法

这种方法是在编制课程标准和考试指导文件中对该课程的学习和考核标准进行明确的界定,规定每一成绩等级的核心内容,并按照这种标准对学生的学业成绩做出等级评定。根据实际情况,成绩可分为五等,也可分为七等。例如,能常由 16 岁左右的英国高中学生参加的'中等教育普通证书(GCSE)'考试,其成绩按照该课程的英国国家标准所规定可分别判定为 A、B、C、D、E、F、G 七个等级。英国 GCSE 考试的国家标准能够较明确地界定一门课程的核心内容及相应的成绩等级标准,评定标准更具体、更稳定。

所谓相对评价等级制是根据学生的卷面成绩在同考生团体中所

处的相对位置来确定成绩等级分数的评分体制。常见的方法有：

1. 五分等级制

具体评定方法是把学生的成绩(通过考试或教师综合评估后得到的成绩)从高分到低分排列,然后根据考生总数按照下面两种方案所给的比例,对属于各个等级的人数进行套评。

a. 正态分布面积比例法

这种方法是根据人的能力通常服从正态方布的假定,把正态分布曲线下位于正负三个标准差范围内的面积以等距的方式划分为五个部分,使这五个部分的面积百分比结构为 3.5:24:45:24:3.5,然后把一批考生的成绩从高到低依次按上述五个部分的面积比例来确定五个等级的人数及比例,即 A 等(5 分)占 3.5%,B 等(4 分)占 24%,C 等(3 分)占 45%,D 等(2 分)占 24%,E 等(1 分)占 3.5%。

b. 按 1:2:4:2:1 的比例分配法

这种方法是 1948 年日本学校教育法中规定的一种学习成绩评定方法,即把评为 A 等(5 分)、B 等(4 分)、C 等(3 分)、D 等(2 分)和 E 等(1 分)这五个等级的人数按照规定的 1:2:4:2:1 的比例进行分类,形成中间多,两头少的人数结构。

2. 十分等级制

这里的十分等级制是根据心理学中的所谓标准十分量表的原理,对学生的成绩做出 10 个等级的相对评价。这 10 个等级从高到低分别用 10 分、9 分、8 分……2 分、1 分来表示,其各个等级的人数比例是根据正态分布的原理进行划分,具体比例如下表所示：

等级分	10	9	8	7	6	5	4	3	2	1
所占人数百分比(%)	2	5	9	15	19	19	15	9	5	2

不难知道,相对评价法的十分等级制不考虑全体学生的整体学习水平,一律把学生的考试成绩从高分到低分按照既定的人数比例来划定等级分数,不利于调动师生教学的积极性。因此,这种相对评

价法在非选拔性考试中现在已用得很少。尽管如此,在许多情况下,人们在使用绝对评价法下的十等级制时,为了适当平衡各个等级分数的人数比例,也常常借鉴相对评价法下各个等级分数的分配比例。事实上,对一些正规的考试(如证书考试),无论是国内还是国外,在学校实行等级制评定成绩时,往往都倾向于把绝对评价法的思想和相对评价法的思想结合在一起,建立既有利于调动师生双方教学活动的积极性又能够形成合理的考试成绩结构的评分制度。

国外的教学检查和评价

教学检查和评价是教学过程的一个不可缺少的环节。它对了解教学效果,改进教学起着十分重要的作用。因此,二次大战以来,国外对教学检查和评价进行了深入的研究,提出了一些新的主张,进行了一些改革,从中我们可以看出一些发展趋势。范树成老师介绍如下:

一、在教学检查方面

(一)从注重检查学生的知识向注重检查学生的能力方向发展。

过去,考试的目的是检查学生掌握知识的情况,因此,考试的题目学生靠死记硬背的知识就可以回答。这在客观上鼓励了学生片面发展机械记忆能力,出现了"高分低能"的现象。为了克服这种偏向,培养学生的能力,近些年来,考试开始由注重检查学生的知识向注重检查学生的能力、由鼓励死记硬背向鼓励发展能力方向发展。例如,英国在纳菲尔德课程出现以后,为了全面测试和培养学生的能力,鼓励教师把培养能力的目的体现在日常教学中,于 60 年代末为选读这种改革课程的学生设置了相应的纳菲尔德统一考试。这种考试采用了许多新的考试方法来全面测试学生的能力,每门科目的"A"级普通教育证书考试包括六张卷,除了采用选择题全面测试学生对考试大纲所规定的基本知识掌握情况以外,还采用了情境问题、阅读理解问题、实验考查和实验课题研究来测试学生应用知识的能力、手脑并用能力以及创造能力。又如 1980 年,英格兰的一个考试委员会的物理考试共分四次进行:第一试,考查实验能力,成绩占 15%;第二试,考查基础知识的掌握情况,成绩占 30%;第三试,考查知识的综合运

需要全本请访问 www.ertongbook.

用能力,成绩占 35% ;第四试,主要考查能力的强弱,成绩占 20%。美国一年一次中学数学考试的命题也很巧妙,有的考查学生抽象思维能力 ;有的以考查理解为主 ;有的综合了多方面的知识内容,考查学生的综合、概括能力 ;还有的考查学生对教学语言的理解和运用。这种考试的评分标准也比较灵活,有助于学生去大胆猜测、判断。

(二)从注重教师检查向注重学生的自我检查,注重对学生自我检查习惯和能力的培养方向发展。

过去,教学检查主要由教师来进行,对学生的自我检查和自我检查习惯与能力的培养是重视不够的。在教学实践中,虽然有个别教师也注意了这个问题,但是,在教学检查的理论中,并没有对这个问题进行研究或研究得不够。二次大战以来,这个问题才引起重视。苏联教育家达尼洛夫和叶希波夫指出,学生自我检查对于学生提高知识质量和学习成绩具有重要作用。他们要求教师要重视学生的自我检查,重视学生自我检查习惯的培养。苏联教育家巴班斯基也十分强调学生的自我检查。他在《论教学过程最优化》一书中指出,学生应该‘对学习效果进行自我分析,把学习效果和自己的可能性作一番比较,评价使用时间的合理性”。为了促使学生进行自我检查,巴班斯基还提出,教师在进行教学检查时,应该对学生自我检查的情况进行检查。目前,苏联教育领导机构已对学生的自我检查提出了具体要求,要求从小学一年级起就要培养学生独立检查的习惯。苏联 1982 年颁发的《技能技巧发展大纲》中对各年级学生的自我检查都提出了明确的要求,例如,对六岁入学的预备班学生提出,学会按样本检查自己的作业,对八年级学生提出进行自我检查、自己评价的学习和自学活动。

美国心理学家布鲁纳也十分重视学生的自我检查和自我检查能力的培养。他指出,在结束某一部分学习时,对于学习的结果有两种不同的对应方式。一种是尽量使学生自己去领悟成功或失败的程度,获得如何改进的信息,这是一种内在性的对策 ;另一种是教师对

学生的作业进行检查,并给以奖励或惩罚,这是一种外部的对策。这两种方式具有不同的功能,前者具有使学生获得信息的功能,后者具有使之减退学习积极性的功能。因此,布鲁纳认为,学生的自我检查远比教师的检查优越。他在《论教学的若干原则》一文中指出:"教学是带有临时性的,因为它的目的在于使学习者或解题人得以'自力更生'。提供任何矫正性指示都会带来危险,学习者很可能永远依赖教师的指正"。

目前,美、英、日等国也都把学生自我检查作为教学检查的一个重要方法。

(三)从传统考试向标准化考试方向发展。

由于标准化考试具有传统考试所不具备的许多优点,如试题取样范围大、效度高;试题答法简单、明确;评分客观、准确等。因此,采用标准化考试的国家越来越多,美、英、西德等国都在不同程度和范围内采用了标准化考试。

(四)从注重对学生学习效果的检查向注重教师教学效果的检查方向发展。

教学检查,顾名思义,包括对学生学习效果和对教师教学效果的检查两个方面。但传统的教学检查及其理论只注重对学生学习效果的检查,而忽视对教师教学效果的检查。虽然学生的学习效果也在一定程度上反映着教师的教学效果,但学生的学习是受多种因素影响的,因此,其学习效果并不能完全反映教师的教学效果。苏联教育家巴班斯基提出应从以下几个方面来检查、分析和评价教师课堂教学的效率:(1)教学的最优组织原则,即看课堂上是否最优地利用基本教学原则,是否把各种教学方法和组织形式最优地结合起来,查明解决各种基本任务(教养、教育和发展任务)的情况;(2)课堂教学的任务和结构;(3)课堂教学的内容;(4)教学方法、方式和手段的选择;(5)培养一般学习技能和技巧的活动;(6)教师所选择的教学形式;(7)教师同学生谈话的风格;(8)课堂教学的卫生条件;(9)课堂教学的效果。

(五)从根据主观经验确定教学检查内容向根据教学目标确定教学检查内容方向发展。

过去,教学检查内容的选择和确定是没有客观标准的,教师常常根据自己的经验把自己认为重要的部分作为教学检查的内容。因此,美、英等国家提出根据教学目标来确定考试的内容。例如,英国有许多人根据美国心理学家布卢姆的认知目标的分类来制定考试的试卷。美国心理学家加涅主张采用标准参照测量来检查教学效果。标准参照测量是从规定学习目标开始的,然后,再规定试题的范围。在确定试题的范围时,要使其准确地反映学习目标所确定的要求。这样,就规定了试题范围的效度。同时,试题范围的确定必须使测验题目的编制符合评定学习重要结果的要求。

(六)从教学检查方法的单一化向教学检查方法的综合化方向发展。

任何一种教学检查方法都不是万能的。因此,许多国家开始重视综合运用各种教学检查方法,以扬长避短,达到教学检查的目的:

(1)把定期考试和日常检查、平时观察、教师检查和学生自检等方法结合起来。例如,苏联在教学检查中,既有定期考试,又有日常检查和观察 既有教师的检查,又有学生自检。英、美、法、日、西德等国也都综合运用多种方法检查学生的学习情况。

(2)把闭卷、开卷、口试和实际操作考试等考试方法结合起来,使考试方法更加灵活多样。例如,英国一些考试委员会和一些学校举行的考试,都非常重视对学生实验能力和实际操作能力的考试。有的学校还采取口试、开卷等方式进行考试,考试时允许学生带笔记,考试时间不加限制,学生可以回家做,也可以互相讨论。苏联也采取类似的办法进行检查。

二、在教学评价方面

(一)从单纯对学生知识片面评价向对学生学业成绩全面评价发展。

过去,无论是传统的考试,还是标准化测验,都强调量化,对于难

以量化的内容就被排斥在检查范围之外。因此,教学评价也只能是对学生掌握知识情况的评价。无疑,这种评价是极为片面的,它不能全面反映学生的学业成绩和促进学生的全面发展。美国心理学家布卢姆 50 年代提出的教育目标分类的理论,对完善教学评价理论和学校的教学评价工作起了重要作用。从此,教学评价开始从单纯对学生知识的评价向对学生学业成绩的全面评价发展。教学评价不仅要评价学生掌握知识的情况,还要评价学生的分析、综合、理解、创造等高层次的能力,评价学生的兴趣、态度、情感等非智力因素。例如,法国、西德、美国对学生学业成绩的评定都包括知识、能力、兴趣、爱好、情感、意志等几个方面。近 20 年来,英国许多教师都希望教学评价能够真正提供下述问题的确实可靠的情况:学生已学到的知识和技能,他们运用这些知识的能力以及他们已具有的想象力、洞察力和创造力如何。

(二)从相对评价向绝对评价方向发展。

二次大战以前,教学评价所采用的方法是相对评价。它是以鉴别学生与学生之间的成绩差异为指导思想的,其评价结果只表示学生在所有学生中所处的位置,并不表示学生的实际水平,不能评价教学目标是否完成、完成得怎样,以及学生的学习成绩与教学目标之间的差异。因此,尽管这种方法实现了高度的数量化,但它存在着许多不足之处。1963 年,格拉泽发表文章,提出在今后的教学评价中,应着重于绝对评价的主张。绝对评价是以评价学生的成绩与教学目标之间的差异为指导思想的,它把学生的成绩与教学目标相比较,提供达到目标情况的信息。这种评价能够诊断学生学习的优缺点,使学生明确今后的努力方向。因而这种评价的理论和实践得到很快的发展。

(三)从总结性评价向形成性评价方向发展。

60 年代中期以前,国外研究和实行的是总结性评价。这种评价是在以某种形式确定教学成果之后,就算完事,并不对以后的教学工作起指导和矫正作用。它的目的是要"选拔适合于教育的儿童",而

不是为了改进教学以适应儿童。1967 年,美国的斯克里文提出了“形成性评价”的概念。形成性评价能够帮助教师了解哪些教学内容已被学生掌握,哪些还未被学生掌握,教学存在什么问题,它把评价的结果作为改进今后的教学计划,开展教学活动的依据。所以,形成性评价可以对今后的教学工作起指导和矫正作用,它是指向未来的。今天,教学评价的目的是要“创造适合于儿童的教育”,是为了改进教学,以最大限度地促进每一个学生知识、能力和个性的发展,因此,形成性评价得到了越来越多的国家的运用。

(四)从教学评价方法的单一化向教学评价方法的综合化方向发展。

国外学者认为,教学评价方法的综合化有助于克服单一的教学评价方法给教学带来的不利影响,因此,主张把各种教学评价方法有机地结合起来:

1. 相对评价和绝对评价相结合

日本学者板仓圣宣提出的到达目标理论就已包含了这种思想的萌芽。他把教育目标分为“方向目标”和“到达目标”,提出对“方向目标”采用相对评价法,对“到达目标”采用绝对评价法。这样,他把“相对评价”和“绝对评价”割裂开来了。日本另一位学者梶田睿一的教学评价理论克服了这一缺点,他把到达目标分成“达成目标”、“提高目标”和“体验目标”三大类,其中,每一类都包含一系列的知、情、意领域的目标。这样,就把“到达目标”与“方向目标”、相对评价和绝对评价结合起来了。现在,一些国家在教学评价实践中,已经把相对评价和绝对评价两种方法结合起来加以运用了。

2. 评分和评语相结合

过去,评定学生学业成绩采用的是分数形式。现在,美、苏、日等一些国家的学者认为,分数有许多弊端:第一,分数过分强调竞争,给学生带来压力,危害学生身心健康;第二,分数虽有促进学习动机的作用,但这种作用是微乎其微的,它使学生把主要精力放在追求分数上,而不是放在学习一些有价值的东西上;第三,分数并不能完全表

明一个学生学习的全部情况,不能反映出学生在某一学科上有什么优点和缺点;第四,分数使有的学校的教学一味追求它的提高,产生教学上的偏差。因此,一些国家的学者主张采用写评语的办法,要求把评分和评语结合起来。并认为,这种方法可能是所有评分制中最公正的一种。近十几年来,苏联一些普通学校正在进行(主要在低年级)以评语取代传统的对学生知识评价的方法——分数的改革实验。他们取消抽象的 5、4、3、2、1 分数,而以评语对学生的学习成绩作出评价。他们认为,评语可以引起学生的动机和兴趣,可以使学生得到具体的帮助。据调查,实行不记分的方法后,师生间的"敌对情绪"消失了,学生变得更活跃,学习更主动,知识面也较宽广了。

3. 定量评价和定性评价相结合

本世纪 40 年代以前的客观测验追求的是客观性和数量化,因此,对学生学业成绩进行的是定量评价。40 年代以后,一些国家开始重视定性评价,特别是 50 年代教育目标分类理论和 60 年代绝对评价理论的提出,使人们更加重视定性评价。出现了把定量评价和定性评价结合起来的趋势。心理学家加涅把学习结果划分为五种主要类型:心智技能、语言信息、运动技能、认知策略、态度。他指出,要根据不同的目的和试题的内容,对某一类型作不同的评价。例如,运动技能评定,如果要测某人打棒球的准确度和速度,就要进行量的评定,如果是为了评价某人在网球中的发球技术,就要进行质的评定。

需要全本请访问 www.ertongbook.