

中学历史教学评价

束鹏芳 著

东北师范大学出版社
长 春

图书在版编目 (CIP) 数据

中学历史教学评价/束鹏芳著. —长春: 东北师范大学出版社, 2005. 1
ISBN 7 - 5602 - 4001 - 1

I. 中... II. 束... III. 历史课—教育评议—初中
IV. G633. 512

中国版本图书馆 CIP 数据核字 (2005) 第 000493 号

责任编辑: 王会杰 封面设计: 李冰彬
责任校对: 伊 然 责任印制: 张文霞

东北师范大学出版社出版发行
长春市人民大街 5268 号 (130024)

销售热线: 0431—5687213

传真: 0431—5691969

网址: <http://www.nenup.com>

电子函件: sdcbs@mail.jl.cn

东北师范大学出版社激光照排中心制版

长春市第九印刷厂印刷

2005 年 1 月第 1 版 2005 年 1 月第 1 次印刷

幅面尺寸: 148mm×210mm 印张: 11 字数: 368 千

印数: 0 001 — 5 000 册

定价: 15.00 元

序 言

我同束鹏芳先生认识是在全国历史教学研究会年会上。参加年会的历史教师很多，但他每次参加年会都提交论文，而且是有分量的论文，这就引起我的注意。渐渐地，我们就比较熟了。我发现他很钻研，对有关历史教学研究上的一些问题有自己的见解，对教育界学术界的不正之风也不愿附和。中国教育学会历史教学研究会改选时，他被推举为学术委员。江苏历史教学界出人才，作为特级教师的束鹏芳是其中的一个。

前不久，束鹏芳来电话，请我为他的新著写个序言，我说对这方面内容没有什么研究，但他还是希望我写。我只好勉为其难。原以为专门谈历史教学评价问题的书，篇幅不会太多，待看到厚厚的书稿，才知道分量不小。书中比较全面地论述了历史教学评价方方面面的问题，从历史教学评价的产生、发展，包括发展的不同阶段，特别是20世纪我国的历史教学评价发展过程和存在的问题，到历史教学评价的价值追求，发展性课程评价观，以及面对课堂教学、考试测量、学生的学习活动等教学实践领域里的几种不同内容的评价，都有比较深入的研究，内容十分丰富。从目前看，国内专门谈历史教学评价问题的书不多见，如他写的如此全面的更是罕见。这部书的出版，是对历史教学研究的新贡献。

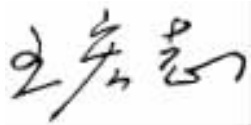
书中展现了历史教学评价问题研究的最新理念与成果。近年来在新的课程理念指导下，历史教学发生了革命性的变化，历史教学评价也应随之变革，不能再停留于纸笔测试认知能力和认知的思维能力上，还应当包括学生的情感态度和价值观，以及历史认知过程、历史学习中的活动表现、历史感受和描述等。因此，要改革历史教学评价的内容和方式方法，要对历史教学评价进行新的探索。他提出一些改革的设想，如应当对开放性考试评价进行积极探索，因为这种考试是从人文关怀目标出发的。书中还展现了近年来历史教学改革的新成果，如，合作学习与探究性学习，作者从人的文化教养和“以人为本”的精神出发，提出了这两种学习的评价指标与方法。这些意见具有时代精神。

在现今的历史教学论著中，有不少是谈理论多，讲实际应用少。

理论固然可以给老师以启迪，但对实际教学的用途不是很大。而这本书不然，不仅有一定的理论高度，而且有很多生动的案例，实用性很强，理性阐述与案例穿插，相得益彰。在历史教学的众多书里，本书的这个特点比较突出。加之作者文学功底好，语言活泼，时常运用生动的比喻，可读性也比较强。本书还在评价的一些理论和概念的梳理上作出了积极的探索和较清晰的脉络梳理。我相信，这也有助于中学历史教学评价在新课改的方向上正确推进。

对书中的一些见解，我很赞同。如，在谈到中国的历史教学评价从总体上滞后于西方时，他主张要向西方学习。在怎样学习的问题上，他引用了加拿大的范梅南教授的说法：东方在向西方学习的过程中，想学习好的地方，但是拿过来以后就变坏了……比如考试，中国古代的考试本来是很大文化的，考艺术，考诗歌，考对策或者哲学，但现在学习西方的标准化考试，而且标准化考试越来越厉害。的确，我们在学习西方的时候，应在充分研究的基础上学习，而不能盲目照搬或加油添醋，更不能把自己过去好的东西完全抛弃。最近我看到1902年10月，京师大学堂举行招生考试的各科试题，其中有一道中国史学题：“问汉武帝盛击匈奴，唐太宗厚遇突厥，御外之术各有不同欤（这里的“外”指的是少数民族）？试略述其得失。”这个题目不是背书本就能回答的，要在掌握一定史实的基础上动脑筋思考才能答出来，而且，还要具有民族情感和正确的民族观，才能答好。我们的前辈在那个时代能出这样的题，很不简单，比起从西方泊来的标准化试题要好得多。当然，这几年的高考试题已经有所改善，标准化的题目减少了。总之，在历史教学领域，我们应继承发扬自家的长处，学习外国先进的东西是很必要的，关键是怎样学。作者在这方面做得比较好，既注意吸收国外历史教学评价的长处，又注意结合本国的历史教学实际，在此基础上提出自己的评价创意，撰写成此书。

祝愿本书早日与读者见面，相信它一定会受到广大教师的欢迎。



2005年元月于沙滩

上篇:出自宏观视野:历史教学评价的支架	1
第一章 中学历史教学评价的历史回顾	2
第一节 20 世纪以来的国内外课程评价改革	2
第二节 20 世纪以来的中学历史教学评价	15
第三节 历史教育的功能与质性评价	25
第二章 历史教学评价的价值追求	38
第一节 知识与能力层面上的价值追求	38
第二节 过程与方法层面上的价值追求	46
第三节 情感态度与价值观层面的价值追求	55
第四节 三维目标上的两个统一	62
第三章 中学历史教学评价的整体框架	71
第一节 中学历史教学的评价原则与评价内容	71
第二节 中学历史教学的评价类型与方法	82
第三节 中学历史教学评价的方式	91
第四节 精神建构:中学历史新课程评价的意义 ..	102
第四章 基于主体性发展的中学历史发展性评价	108
第一节 发展性评价概述:理念与特征	108
第二节 发展性学生评价	113
第三节 发展性教师评价	130

下篇:立足于特定场域:教学评价的抛锚	151
第五章 基于过程和情境活动的表现性评价	152
第一节 表现性评价概述	152
第二节 质性评定:成长记录袋的评价方法	163
第三节 质性评定:苏格拉底式研讨评定法	180
第六章 侧重认知与认知思维测量的考试评价	198
第一节 考试评价及其改革趋向	198
第二节 考试评价的测量目标	212
第三节 考试评价的命题新要素	224
第四节 开放式考试评价	242
第七章 描述课堂教学的形成性评价	262
第一节 一堂历史好课的理性根据与实证分析	262
第二节 课堂教学评价的维度与量表	280
第三节 教学进行时:特定环节与特定视角的 专项评价	293
第八章 合作与探究“场域”里的质性评价	313
第一节 历史教学中的合作学习	313
第二节 历史教学中的探究性学习	326
参考文献	342
后 记	344

上篇

● 出自宏观视野：历史教学评价的支架

- ◇ 中学历史教学评价的历史回顾
- ◇ 历史教学评价的价值追求
- ◇ 中学历史教学评价的整体框架
- ◇ 基于主体性发展的中学历史发展性评价

第一章 中学历史教学评价的历史回顾

第一节 20 世纪以来的国内外课程评价改革

一、课程评价活动的出现

有教育就会有评价，早期的教育评价是教师对学生个体的口试，也就是进行口头测试，在学生数量增多的情况下，就随之改用笔试，这是课程评价的雏形。较正规的课程评价的“渊源”应该可以追溯到我国隋朝出现的科举考试，距今已有 1400 多年的历史；从教育自身发展来着眼的系统评价活动则是 19 世纪末出现在美国的记忆状况测定，距今才 100 多年。因此，课程评价是一个既古老又年轻的领域。1897—1898 年，美国学者莱斯（T. M. Rice）曾对 3 万多名小学生进行拼字测验，以检验拼字教学时间对学习效果的影响，评价界一般以此作为评价研究的开端。课程评价作为一门学科，在短短的百年时间内，理念几经变换，思想不断更新。不同的学者根据各自的标准进行了不同的划分，提出了不同的见解。

二、国外课程评价的历史发展：时期与特征

以色列著名教育学家、课程论专家利维（A. Lewy）对教育评价的历史进行了大跨度的审视，将其划分为三个时期：

（1）古典的考试型时期（examination pattern）。教师主要以口头提问的形式检查学生是否已理解所学知识并记住一些重要部分。如西方中世纪时期的宗教教育以及我国传统的私塾教学，使用的多是这种考试类型。

（2）心理测量占统治地位的时期。它始于第一次世界大战后的

20 年代，是科学管理效率测量工具在学校成绩测验中的应用，形式大多是多项选择的纸笔测验，这种测验形式直到 20 世纪 80 年代末，仍在学业成就评价中发挥着极大作用。

(3) 后现代时期 (postmodern era)。它兴起于 20 世纪 80 年代以后，其最大特征有二：一是增加了教师在评价领域中的权威，二是对某些心理测量原则的合适性质疑，并以开放性的结论以及“评定”(assessment) 这一概念代替以前的“评价”(evaluation)。

利维指出，“评定”是 20 世纪 80 年代以来流行于评价领域的术语，有学者从词源学的角度分析，它的词根是拉丁动词“assidere”，意即作为评价者的教师“坐在考生旁边”。这个词暗含着这样的意义：评价者由教师来担当，但他的行为不仅是评价者，而且是为了学生的利益并支持学生。

美国评价专家古巴和林肯 (E. G. Guba & Y. S. Lincoln) 对评价领域的划分以及对该领域的批判则深刻地把握了评价发展的内蕴。这里，我们试以他们的划分和批判为线索，探究评价理念的发展与变化。

(一) 测验和测量时期

以各种形式对学生进行测验的课程评价产生于 19 世纪末，盛行于 20 世纪 30 年代。由于自然科学的发展，各种统计、测量技术得到长足发展，一批教育、心理学家开始把这些技术运用于教育领域，如英国遗传学家、心理学家高尔顿的《英国科学家：他们的秉赋与教养》、法国的“比奈—西蒙智力量表”、德国冯特的实验心理研究成果和美国桑代克的《心理及社会测量理论》都相继问世，各种智力的、学业成就的、人格的测验工具也随之涌现。这一切都为评价的正规化、系统化创造了条件。同时，“一战”后发展到顶点的工商业的“科学管理运动”也对学校教育发生了深刻影响，“学校被视为工厂”、“学生被视为原料和产品”、“教师成为加工者”。学生产品是否符合需要，教师教学的成效如何，学校教育成功与否，似乎都可以通过“测量”来检验。这显然为评价的发展提供了客观的社会需要。评价由此而得以兴盛，课程评价史上的第一代评价也因此形成。第一代评价的

基本观点是：评价在本质上是以测验（testing）或测量（measurement）的方式测定学生对知识的记忆状况或某项特质。因此，评价就是测量，评价者的工作就是测量技术员的工作——选择测量工具、组织测量、提供测量数据。这一时期被称作“测验”和“测量”的时期。

（二）描述时期

20世纪30年代，美国完成初等教育的人数激增，但经济大萧条又使大批青年找不到工作而无处可去，只好到中学注册学习。中学教育因而急剧膨胀，原有的中等教育目标、课程、评价标准都受到前所未有的挑战。由进步教育协会主持的“八年研究”，就是试图面对这些挑战，回答如下问题：中学除了学术性课程是否还应增加其他课程，从而适应那些完成中等教育后并不打算进一步接受高等教育的学生的需要？增加了其他课程是否会降低教育质量进而影响高等教育的质量？等等。作为一项课程实验，“八年研究”无疑是成功的，并且它还特别推动了课程评价领域的发展。“以前，如果对教育计划或程序的结果进行客观评估，是根据学生在成绩测验中的表现作出判断，而成绩测验通常是测量简单的技能和对信息的回忆，并用一个分数概括地报告结果。八年研究提醒教育者，教师在一项课程实施中通常要求几项教育目标。研究还表明，通过使用问卷、观察、产品样本和测验，都可以评估学生在每个主要目标上的进展情况。这已成了一项普遍的评价观。”“八年研究”为新课程实验设计了一套课程评价办法，被称为“划时代的课程评价宣言”。作为“八年研究”评价组主任的泰勒曾指出，评价应该是一个过程，而不仅仅是一二个测验。评价过程中不仅要报告学生的成绩，更要描述教育结果与教育目标的一致程度，从而发现问题，改进课程教材和教育教学方案方法。其课程评价的基本思想可以概括如下：（1）课程评价就是确定具体课程目标实际达到的程度；（2）对人的行为评价应该是多方面的综合；（3）评价方法应该是多样的，不能局限于某种评价方式。他们强调评价要以课程目标为核心，以全面发展人的才能为主要目标，这就形成了目标导向评价。因而教育目标就成了评价能否展开的关键。

在泰勒的影响下，一时间，美、英等国出现了诸多针对评价而设计的教育目标体系，其中以泰勒亲自参与并指导的布卢姆的教育目标分类学影响最为深广。布卢姆等人明确提出，制定教育目标是为了便于客观地评价，而不是表述理想的愿望。事实上，只有具体的、外显的行为目标才是可测量的。布卢姆等人曾以公式来表示，这就是“目标=行为=评价技术=测验问题”。从此，泰勒原理与布卢姆的教育目标分类学一起广为传播，成为一项国际性的教育原理。以美国教育心理学家布卢姆为代表，他们把教育目标分类成认知、情感和精神运动三个领域，每个领域在实现最终目标的过程中，各自建立了相应的目标系列。例如认知领域，有六个由低到高的递进层次：知识、理解、应用、分析、综合和评价。这在以后的章节里将作具体的解释。

泰勒的成就和观点影响了整整一代人，在评价领域产生了巨大影响，形成了一个以“描述”为标志的评价时代。这被称为第二代评价，从20世纪30年代一直持续到50年代。这一代评价的核心观点是：评价在本质上是“描述”（description）——描述教育结果与教育目标相一致的程度。教育评价专家克龙巴赫（L. J. Cronbach）写道：“1935年以前，对学生的考查基本上都是依据事实性知识和掌握基本技能的情况。是泰勒在这一时期的研究和论述，才使人们意识到：通过简单的事实测验不能引发高级心理过程；促进事实性知识的教学，并不可能促进——事实上可能干扰——其他重要的教育目标。”也正是在泰勒的努力下，“评价”（evaluation）一词取代了“考试”（examining）和“测验”（testing）两个术语，使评价领域有了一种新的指导思想。总之，第二代评价的基本特点是：认为评价过程是将教育结果与预定教育目标相对照的过程，是根据预定教育目标对教育结果进行客观描述的过程；评价的关键是确定清晰的、可操作的行为目标；评价不等于“考试”和“测验”，尽管“考试”“测验”可以成为评价的一部分。第二代评价和第一代评价相比，使评价走上了科学化的历程，但是这一代目标评价特别强调行为目标，清晰的目标既是评价的起点又是评价的终点，这就使课程计划陷入一种预测模式之中，而在预测以外的有价值的东西以及目标以外的课程学习结果却被

忽略了。

（三）判断时期

1957年以后，美国因苏联卫星上天而发动了新一轮的教育改革。在这一轮教育改革浪潮中，评价人员开始关心一个问题，即：对已经确定的目标，是否需要评价，是否需要价值判断？由此延伸开去，评价专家又激烈地讨论了判断是否应成为评价的一项基本活动，判断是否需要标准，如果需要标准，能否建立科学、客观的“价值中立”的标准。通过讨论，人们意识到课程目标作为评估活动的依据不是唯一的，目标本身存在的科学性、合理性和可行性也应该进行检查。教育是个人自我实现的过程，用统一的目标模式去评估每个人的自由发展，去评估教学效果，从根本上说是不可接受的。因此，以课程目标为中心的评价模式受到了强烈的冲击。斯太克认为，传统的评价模式是“预定式的评价”（preordinate evaluation），即先陈述目标，再依目标来搜集资料，并对目标与结果之间的差距或一致程度作出报告。这种评价的缺陷是，对课程方案本身的合理程度、变化、学生在目标之外的感受，各方面人士对目标、课程实施等方面的不同观点都难以反映，而且教育的价值，有时还是扩散的、潜在的，有时又是内在的，这些也都难以用事先确定的目标校对。斯太克还认为，评价的意义在于服务；为了使评价有利于服务对象，评价者就应该首先关注服务对象关注的问题、兴趣和焦点。因此，“回应”服务对象是评价的起点。艾斯纳认为：“凡是内容、活动、目标、顺序或呈现方式、反应方式，都必须考虑各种抉择，评价各种计划的用途评价就是要对某些计划方案的优劣进行价值评估。评价的主要问题，在于挑选时的集中性、复杂性和综合性。”斯克瑞文也认为评价是一种方法上的活动，它是根据一组加权（weighted）的目标标准，收集和综合有关的现象资料，以形成比较的或价值的判断，并说明资料收集方式、加权数和目标选择的合理程度。第三代评价的基本特点是：把评价视为价值判断的过程，评价不只是根据预定目标对结果的描述，预定目标本身也需要进行价值判断；既然目标并非评价的固定不变的标准，那么评价就应当走出预定目标的限制，过程本身的价值也应当是评价的有

机构成。评价在本质上是“判断”（judgement）。这样看来，第三代评价是对第二代评价的重要超越，它走出了第二代评价“价值中立性”的误区，确认了价值判断是评价的本质，确认了评价的过程性。许多新的评价理念如“形成性评价”、“目标游离评价”、“内在评价”等都在这一阶段产生。从20世纪50年代中期持续到20世纪70年代的评价研究和实践，被认为是课程评价的判断时期，它推动评价向着过程性、多元化发展。判断时期的代表人物提出了反映各自价值取向的评价模式，比较有代表性的评价模式是“CIPP”模式，即背景 Context 评价、输入 Input 评价、过程 Process 评价和成果 Product 评价融合为一个完整的评价系统。它把过程改良视为评价的首要任务和基本特征，从而摆脱了评价的控制性。

（四）建构时期

20世纪60年代末、70年代初，随着对课程改革运动的深刻反省，传统的评价方式也受到猛烈的冲击，人们渴望发展评价的新理论和新范式。1972年11月，在英国剑桥大学的邱吉尔学院（Churchill College），14位学者会聚一堂，讨论评价理论和范式的创新问题。会议建议采用新的、文化人类学的研究范式取代旧有范式，即：评价不是对预期的教育结果进行测量的模式，而是要对整个方案，包括前假设、理论推演、实施效果以及困难问题等等进行全面而深入的研究，质性研究范式由此推广开来，并开创了一代新的评价理念，即古巴和林肯所谓的第四代评价。

古巴和林肯认真考察前三代的评价理论，在肯定成绩的同时，指出了它们存在的严重缺陷和问题。

一是“管理主义倾向”。教育管理者总是控制着评价，决定着评价的范围和任务，决定着评价的对象，而管理者却处在评价范围之外。这种关系在实际活动中常常造成不合理的后果：管理者无过失，管理者与评价者的关系有失公平，决策权掌握在管理者手中；不利于被评价者按评价建议采取有效的改进行动，导致一种有害的默契：管理者用各种办法保护自己不受损害，评价者则保证使用管理者认可的方法，以换取合同的兑现，保证评价的经济收入。

二是“忽视价值的多元性”。“社会在根本上是价值多元化的”，这种多元性不仅表现在政党、利益集团之间，也深刻反映在个人的生活中，反映在人们对待许多具体的问题上。由于价值多元性的存在，人们常常要问：“这一教育评价是谁作的？”“为谁作的评价？”所以，受到伤害的被评价者就会采取不合作态度，“客观的”评价结果也就难以被具有其他文化背景和其他价值观念的人接受，于是，希望通过评价来改进课程、教学、教育的期望也就落空了。

三是“过分依赖科学范式”。这种范式的问题在于：打着“科学是纯粹价值中立的”旗号回避价值问题，逃避评价者的道德职责，以图寻求评价的客观性和中立性。但是，“人类的所有活动，多少都具有价值的成分，即使是自然科学也不例外”，过分依赖实证科学范式，使评价过于依赖“数的测量”而忽视“质”的探究；过分依赖实证科学范式，使评价者忘记了“科学方法”、“实证技术”只是人类认识、评价事物的一类方法，而远不是全部方法。正是实证科学的范式在“强化和支持着管理主义”。“数据资料自己不会说话”，在管理主义支配下，资料数据虽然是评价者用实证科学、数量化的方式获得的，但这些数据资料却是评价者首先按照管理者规定的要求和范围，然后再用实证方法去提取的。这一点却常常为人们忽视。

古巴和林肯在回顾和批判的基础上，提出了他们所谓的“第四代评价”（fourth generation evaluation）。第四代评价的中心思想是，评价在本质上是一种通过“协商”而形成的“心理建构”，因此，评价应坚持“价值多元性”的信念，反对“管理主义倾向”。在具体方法上，第四代评价更强调在自然环境中，用质性研究方法，使各方人士通过各种形式的对话达成共识。在此特别要指出的是，第四代评价突出了过去作为被评价者的学生在评价过程中的参与者身份，从而使得评价中出现了被评价者的声音。评价者在协商过程中没有特权，不应采取任何控制他人的态度，而应该是一个中介人、一个条件提供者、创造者。协商达成的共识，首先是有关各方的共识，评价者只是其中的一个方面。

第四代评价的基本特点是：把评价视为评价者和被评价者“协

商”进行的共同心理建构过程；评价是受“多元主义”价值观所支配的；评价是一种民主协商、主体参与的过程，而非评价者对被评价者的控制过程，学生（被评价者）也是评价的参与者、评价的主体；评价的基本方法是质性研究方法。

第四代评价旗帜鲜明地突出了评价中的价值问题，从而突破了评价领域中长期以来所寻求的“客观性”、“科学性”迷雾，使评价的理念发生了质的飞跃。在过去的一个世纪中，大多数评价专家所追求的是评价在过程、手段以及结果上的客观科学化，这在某一个方面是有其积极意义的，但把它强调到极端，就会导致忽略评价的价值特性，造成评价中的不合理现象，以致人们对评价失去信心。第四代评价意识到了这一缺陷，于是重新回到起点，首先探讨“评价是什么”，在确认评价的价值本质的基础上，再来探讨评价的方法、过程和结果，这从深层次促进了评价理论的发展。在此基础上，它所倡导的“协商”式的“共同心理建构”，实质上是尊重每一个个体的主体性，并在此前提下寻求共识的达成，这反映着一种深刻的民主意识，极其富于时代精神。

当然，正如它的倡导者所说，“第四代评价也只是一种建构”，它本身并不完全排斥其他的评价模式，而应该视具体的评价任务，与其他的评价模式相互补充。事实上，第四代评价的提出，是对当前评价领域发展趋势的一种概括和总结，在它的旗帜下，可以涵盖许多新产生的评价模式和方法，如“教育鉴赏与教育批评”（education connoisseurship and educational criticism）模式，产生于教育实践中的“档案袋评定”（portfolio assessment）及“苏格拉底式研讨法”（Socratic seminars）等，都与古巴和林肯所谓的第四代评价有着相同的理念和追求。目前，这种评价思想正为越来越多的人所接受，且在评价实践中得到广泛应用。质性评价范式的应用标志着建构时期的到来。

西方案例 日记里的评价

又到了写日记的时间，老师和他们的孩子每周一次的早晨花上20分钟时间写日记。尽管老师没有要求孩子们这样做，但是，他们

还是经常将他们的日记放在老师的桌上。老师看完后，还给学生，并加以口头评价。孩子们和老师都非常珍惜这段时间，它给大家创造了反思和表达用其他形式不容易表达出来的感情的机会。……每一个孩子都有一个故事，这个孩子在学校的表现必须依据这个故事才能加以测量和理解。当老师正在自己的日记本上写她自己家里的两个孩子的事的时候，她抬起头来的一刹那，目光停在了尼可尔的身上。尼可尔的父母合不来，母亲和女儿生活在一个凶暴的丈夫和爸爸的恐惧之中。……老师问道：“没有什么事吧，尼可尔？”尼可尔回答说：“没有。”但是，老师知道一定出了什么事，她用慈善的目光关切地看着她：“真的没有什么事吧，尼可尔？”在理解的目光下，尼可尔找到了表达他的需求的可能性：“如果爸爸知道我们要搬走的话，他会杀死我们的。可是我会想念我的爸爸的，我还会想念我的老师，她对我来说，太重要了。”她写道……老师看到尼可尔的脆弱，也看到了尼可尔的成长期待。^①

简评：在孩子的故事里给予理解和评价，这就是一种质性评价，是将评价与教育过程贯穿在学生的成长之中的。

三、我国课程评价的历史发展：独立与追随

我国历来是一个教育大国，许多现代的课程思想包括课程评价思想都能够在古代找到其最初的萌芽。战国时代的《礼记·学记》中提到了对学生管理和考核的规定与要求，如：“比年入学，中年考校：一年视离经辨志，三年视敬业乐群，五年视博习亲师，七年视论学取友，谓之小成。九年知类通达，强立而不反，谓之大成。”它们明确规定了先考什么、后考什么、学习年限以及规格要求等，既反映了我国早期的学校体系已经相当完备，又体现了课程评价与学校教育如影相随的规律。隋朝的科举考试分科选拔，逐级考试，并在考试方法、成绩发布、录用程序等方面形成了较完备的制度，这是较系统的课程

^① [加拿大] 马克斯·范梅南. 教学机智：教育智慧的意蕴. 李树英译. 北京：教育科学出版社，2001

评价活动的萌芽。唐代分科更细，宋朝还采用了糊名法，课程评价的价值趋向于对客观公正的追求。到明清时期，科举考试转变为四书五经范围内命题的单一科目的八股考试，在八股考试的评价引领下，考什么学什么、怎么考怎么学的现象日益显著，类似于现在的指导升学考试的“时文”、范文十分畅销，学业评价唯以考取与否为标志。僵化的学业评价未能使中国的课程评价思想和实践与时俱进。

20 世纪初期，我国废除科举制度，值癸卯学制确立时，西方课程测验运动正方兴未艾，于是中国的课程评价开始了对西方理论的追随，20 世纪 20—30 年代形成了中国的课程测量运动。

1918 年美国瓦科特在北京清华学校任教时，用比纳量表测量了该校高等科四年级学生。1920 年，陈鹤琴等人在南京高师开设测验课程，并用心量表测量了学生，这算是在我国正式开始应用科学心理测验。1922 年美国课程测量学家麦柯尔（Mecall）应“中华课程改进社”之聘来华，将 TBCF 测量编制法介绍到中国。从“五四”前后到 20 年代末，我国的课程测验运动出现了高潮。因为人们赶时髦，认为什么东西都是可测的，以至我国的课程测验出现了偏差和混乱状态。著名学者陆志伟、陈鹤琴等人倡议成立了“中国测验学会”，这是我国课程测量和评价方面的第一个学术研究组织，使测验和测量又继续发展。但是 30 年代中国的民族危机日益严重，在日本帝国主义的大举进攻下，中国的学校教育受到严重破坏，泰勒的“八年研究”（1934—1942 年）及其成果自然就不可能介绍到中国，我国的课程评价自此与世界拉开了距离。

新中国成立后，我国又照搬苏联的课程模式，全盘否定西方国家的课程甚至是旧中国的课程理念。源自一种片面，我们丧失了课程和课程评价的意识，简单地套用以五级分制为核心的苏式成绩考评法。一直到 60 年代，在中苏关系恶化的背景下，苏式考评法也遭到了批判。但是我们没有找到新的课程评价发展的方向，于是，我国的课程评价遂陷于无所适从的困境。

1978 年党的十一届三中全会以后，尤其是高考的恢复，我国的课程评价研究又迅速发展起来，国外及其我国台湾地区有关课程评价