

聚焦新课程系列丛书



JJXKCXLCS

夏谷鸣 任美琴 周 瑜 主编

初中英语 教学评价



CHUZHONG YINGYU
JIAOXUE PINGJIA

东北师范大学出版社

· 聚焦新课程系列丛书



JJXKCXLCS

夏谷鸣 任美琴 周 瑜 主 编

初 中 英 语

教 学 评 价



CHUZHONG YINGYU
JIAOXUE PINGJIA

东北师范大学出版社
长 春

图书在版编目 (CIP) 数据

初中英语教学评价/夏谷鸣、任美琴、周瑜主编.
—长春: 东北师范大学出版社, 2005. 2
ISBN 7 - 5602 - 4003 - 8

I. 初... II. ①夏... ②任... ③周... III. 初中英语
教学评价 IV. G633.412

中国版本图书馆 CIP 数据核字 (2005) 第 013755 号

☐责任编辑: 徐华娟 ☐封面设计: 李冰彬
☐责任校对: 王宏志 ☐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)

销售热线电话: 0431—5687213

传真: 0431—5691969

网址: <http://www.nenup.com>

电子函件: sdcbbs@mail.jl.cn

东北师范大学出版社激光照排中心制版

长春市新世纪印业有限公司印刷

2005 年 2 月第 1 版 2005 年 2 月第 1 次印刷

幅面尺寸: 148mm×210mm 印张: 7.5 字数: 159 千

印数: 0 001 — 5 000 册

定价: 10.00 元

目 录

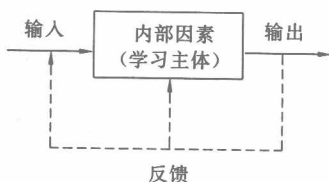
第一章 课程评价概念.....	1
第一节 评价的定义.....	2
第二节 评价的功能.....	9
第三节 评价的主体	16
第四节 评价的种类	20
第二章 形成性评价	30
第一节 形成性评价的功能	30
第二节 形成性评价的原则	35
第三节 需要分析	40
第四节 档案分析	55
第五节 行为评价	71
第六节 访谈评价	85
第三章 终结性评价.....	100
第一节 口语测试.....	105
第二节 听力测试.....	112
第三节 书面测试.....	118
第四节 阅读能力测试.....	128
第五节 对终结性评价的评估.....	145

第四章 书面考试评价样章及评价表·····	148
(一)·····	148
(二)·····	166
(三)·····	184
(四)·····	215
(五)·····	224
(六)·····	232
后 记·····	236

第一章 课程评价概论

- ◆ 了解课程评价的简单发展历史，正确理解课程评价。
- ◆ 全面理解课程评价的三大主要功能，以及它们之间的关系。
- ◆ 反思你如何分析考试试卷，包括与学生交流、与家长沟通情况。
- ◆ 谈谈你对建立评价资源库的看法和打算。
- ◆ 简单陈述你所知道的各种评价手段。
- ◆ 整理你在平时课堂教学中所采用的各种评价方法，并根据你的实践分析它们各自的特点和弊端。

语言教学是围绕学习主体（内部因素）展开的一系列输入（input）和输出（output）活动。从理论上讲，输入的内容和策略应根据输入的对象而定。可是，教学活动中的输入对象往往是一个“黑箱”（black



语言教学图示

box) 或“灰箱”（gray box）。我们只有通过分析外部因素，

尤其是输出端的信号，才能观察其内部结构。这样，整个语言教学活动就构成了一个系统（见上页图）。在这个系统中，输出端（即评价测试）是一个十分关键的环节，它既是一个终结性环节（赋值功能），又为下一个循环提供反馈信息（导向功能）。它不但能帮助系统外部了解内部因素的情况，为调整输入的质量和策略提供依据，而且有助于系统内部完善监控机制（monitor），为自我调控提供反馈信息，刺激内在化活动（internalization），促进中介语（inter-language）的发展。

历次教学改革最后都归结到评价上，评价往往被视作改革的瓶颈。这一方面说明了评价在教学改革和发展过程中的重要地位，另一方面也反映了人们对评价的误解——以为只有改变考试方式，教学改革才能顺利进行。这种误解把评价的功能过分地放大，使评价成为整个教学活动的中心，其实质还是局限于应试教育的意识之中。要走出误区，我们须要认真解读课程评价的内涵和外延。

第一节 评价的定义

什么是课程评价？这是课程领域中一个十分普通而又复杂的问题。任何课程理论流派在构建自己的理论体系时都会讨论这个问题；任何课程专家在回答这个问题时都有符合自己的理论系统的诠释。因此，批判分析各种课程评价的定义，不但令我们对概念本身具有更加清晰的认识，而且使我们能够从中考察课程评价理论及实践的大致发展轨迹。

历史上一出现按一定体系和计划来实施的教育，就有了

教育评价。我国最早的一部教育专著《礼记·学记》（写于战国时期）中就有关于当时考核制度的论述。隋代确立的科举考试制度可看作有系统的教育评价活动的萌芽，在中国延续了1300多年，对世界各国的考试制度产生过深远的影响。

国外早期的教育评价活动主要是考试，有口试和笔试两种形式。美国19世纪后半期多采用口试法。随着教育的发展，学生的增多，口试很快为笔试所取代。在西方，笔试最早是1702年于英国剑桥大学开始使用的，1845年由美国教育家贺拉斯·曼（Horace Mann）在波士顿文法学校首次引进，并逐渐在美国推广和发展起来。但此时的考试仍是论文提交的形式，评分有很大的主观性。为使考试客观化，教育测验运动由此兴起。现代意义上的教育评价产生在20世纪30年代的美国，由20世纪初的教育测量（又称测验、测定）运动发展而来。

如果说在20世纪30年代以前称为前系统评价阶段，那么此后就是系统评价阶段，其明显的标志是美国著名心理学家拉尔夫·泰勒（Tyler, R. W.）提出的泰勒模式。整个阶段分“泰勒时期”（1930—1945）、“萌芽时期”（1946—1957）、“发展时期”（1958—1972）和“专业化时期”（1973— ），至今只有半个多世纪。但作为一门系统的新兴专业领域，课程评价研究发展极其迅速，在短短的半个多世纪中，就已经结出了丰硕的果实。理论家们从不同的角度对课程评价进行界定并构建自己的评价模式，令人目不暇接。在众多观点中，李子建、黄显华（1994）罗列了五种比较有代表性的界定：

1. 将教育或课程评价等同于教育测量和测验

19 世纪下半叶和 20 世纪初的 30 年间, 学者们着重于学力测验的研究和教育成就的定量化、客观化和标准化。在这种形势下, 人们普遍认为评价即测量 (measure)、测验 (test), 其代表人物有桑代克 (Thorndike)、哈根 (Hagen) 和美国教育测验的权威人物伍德 (Wood, B.) 等。的确, 测量或测验一直是评价的一个重要组成部分。美国有位大学教授曾作过的一个调查很能说明问题。他要求一些师范教师和学生立即回答一个问题: “当听到‘评价’这个词的时候, 你首先想到的是什么?” 调查结果按人数递减的排列顺序是:

教师: (1) 测验; (2) 测量; (3) 等级; (4) 绩效; (5) 对隐私的侵犯。

学生: (1) 测验; (2) 等级; (3) 成绩; (4) 不公平; (5) 判断。

教育测量的研究取得了一系列的成果, 在考试的定量化、客观化与标准化方面取得了重要的进展。但是, 当时的考试与测验只要求学生记诵教材的知识内容, 较为片面, 无法真正反映学生的学习过程。

2. 将课程评价视为学生成就、表现和某种特定目标间的符合程度

在 1933—1941 年, 泰勒及其同事开展了一场规模浩大的“八年研究”, 对高中生进行追踪调查, 以改变当时以升学考试为取向的片面做法。研究发现, 其时的课程和测验可以称为“教科书中心主义”。课程是以知识为中心, 教学只要求记诵教材的知识内容, 对学生的评价是测量学生对事实和知识的回忆与认识。另外, 当时测验的性质是常模参照测

验，目的在于按分数高低对学生进行排位和区分，为升学服务，而无视学生的学业进展。

以“八年研究”和对当时测验观的批判为基础，泰勒对课程评价作出了新的界定，即将教育或课程评价界定为学生学业表现和某种特定目标间的符合程度。这种观点具体反映在泰勒原理或称泰勒评价模式中。泰勒评价模式的起源可以追溯到泰勒在 20 世纪 30—40 年代的早期著作，系统阐述是在 1950 年出版的《课程与教学基本原理》。泰勒认为，课程编制须回答四个基本的问题：

- (1) 学校应该试图达到什么样的教育目的？
- (2) 可以提供什么样的教育经验才能达到教育目的？
- (3) 这些教育经验应该如何有效地组织起来？
- (4) 如何确定这些教育目的是否达到？

上述四个问题又可以概括为三个要素：目标、学习经验（的选择和组织）以及评价。所谓评价，就是确定学习结果与教育目的之间的达成度。

泰勒对于课程评价的重新界定和评价模式的提出，标志着课程评价作为一个独立的研究领域诞生了。泰勒的评价观和以前那种把评价与测验混为一谈的做法形成了鲜明的对比，两者在性质、方法、对象和功能上都有很大区别。以前的测试只是一种常模参照测验，运用分数、平均分等简单的做法考查学生的知识，把学生分成各种成绩等第。泰勒评价则是一种目标参照测验，要求预先设定目标，根据目标评定学生的表现。评价方法多样化，包括笔试、观察访问、面谈、问卷调查、收集学生实际作品等。以综合性的目标为参照点，评价的对象不仅是学生学了多少知识，而且评价学生

的智能、体能和品德。评价的功能在于了解学生对学习目标的达成度，从而改进学生的学习或者修正学习目标和内容。

泰勒评价模式一问世，就迅速风靡美国和西欧，掀起了教育评价的热潮。但是，随着时间的推移，人们越来越发现泰勒评价观的一些内在缺陷。人们开始批判泰勒模式的缺陷，酝酿和发展新的评价理论，最后把评价推进到专业化水平。马道斯等（1983）把1930—1945年称为“泰勒时期”。

3. 将评价视为搜集与提供资料，让决策人员从事有效决策的过程

克龙巴赫（Cronbach, L. J.）在其1963年著述的《通过评价改进教程》中，把评价广义地定义为“为作出关于教育方案的决策收集和使用信息”。所谓“方案”，“可以是关于遍布于全国的教材方案，也可以是关于某一个学校的教学活动的方案，或者是关于某一个学生的教育经历的方案”。所谓“决策”，包括教程的改进、关于个体的决策（学生的需求、长处和缺陷等）以及行政的调控（对学校与教师等的调控）。克龙巴赫把评价的功能界定为决策，具有深刻的用意。因为泰勒模式虽然也强调对目标、内容及其教学的改进，但是模式的本质在于把学习的结果与目标相比，只重结果，不重过程。

另一位学者斯塔弗尔比姆（Stufflebeam, D. L.）继承了克龙巴赫的观点，明确指出：“评价最重要的意图不是为了证明（prove），而是为了改进（improve）。”他在1985年与人合著的《系统的评价》中详细地阐述道：“评价是一种划定、获取和提供叙述性和判断性信息的过程。这些信息涉及研究对象的目标、设计、实施和影响的价值及优缺点，以

便指导如何决策，满足教学效能核定的需要，并增加对研究对象的了解。”该定义有三个要点：其一，从评价的性质来看，评价是一种事实叙述和价值判断的过程；其二，从评价范围来看，涉及研究对象的目标、设计、实施和影响等；其三，从评价的功能来看，旨在决策和改进。

斯塔弗尔比姆还进一步设计了方案评价模式。第一步为背景（context）评价，考察现存的目的与重点是否与使用者的需要相一致，目的本身得到了评价；第二步是输入（input）评价，看方案是否可行，包括能否达成目标、成本如何、资源配置情况等；第三步为过程（process）评价，检查方案的实施，提供反馈信息；第四步是成果（product）评价。该模式的另一个版本是CSE评价模式。CSE评价模式因美国加利福尼亚州州立大学评价研究中心（Center for study of Evaluation）的研究和推广而得名。它把上述的四个步骤简化命名为需要评定、方案计划、形成性评价和总结性评价。

总之，这种课程评价的定义由于强调评价的决策、改进功能，强调过程，而大大地拓宽了课程评价的范围。泰勒模式只要求结果与目标的对比，斯塔弗尔比姆的CIPP模式则对课程编制、实施的全流程进行评价。

4. 将评价视为专业人员的判断，对优点和价值的评估

与第三种课程评价定义侧重于评价的功能不同，第四种评价定义侧重于评价的本质属性，认为评价的本质为价值判断，并要求由专业人员作为评价者，其代表人物是艾斯纳（Eiser）。艾斯纳的评价模式名为“教育鉴赏及教育批评取向”。“教育鉴赏是指欣赏细致教育现象及其重要素质的能

力，教育批评则指展示鉴赏结果的方法（通常以批判论文形式表达）”。教育批评包括三个阶段，即描述课堂事件、诠释教育事件和对所描述、诠释的现象作出优缺点的价值判断。

评价究竟是一种客观的事实描述，还是带有主观色彩的价值判断，这本来构成不了一个值得广泛争议的问题。无论是中文还是英文，评价词义都含有价值判断的意思。只不过一些理论家们有意无意地为了某种功利，才否定评价的判断属性罢了。陈玉琨指出其中的两条原因：“其一是试图淡化评价的判断属性，以获取人们对评价的肯定态度。其二是反对统一的判断标准。”内伏（Nevo）指出：“以这种忽略评价的主要特征来引起对评价的肯定态度，也许是不现实的。发展对评价肯定态度的另一途径，也许是在教育的各个领域中证实评价的建设性功能。”由此可见，艾斯纳强调评价的判断属性，可以说抓住了评价的本质。

不过，艾斯纳仅仅要求由专家担当评价者，这是否走到了事物的另一个极端呢？罗伯特·斯坦克（Robert Stake）也赞同价值判断的观点。不过他建议评价者应该以从社区中适当的外来集体或个人那里搜集来的相关资料为指导。罗伯特·伍尔夫（Robert Wolf）则怀疑某个评价者能否公正地判断各方面的观点，故主张使用一种“仲裁模式”。总之，评价具备事实描述和价值判断双重属性，评价者也应该多元化，绝不仅仅局限于专家。

5. 将评价视为一种政治活动，不仅检视课程的效率和管理课题，也理解评价所涉及的道德及美学涵义，并探讨谁会从中得益。

持这种观点的代表人物贝蒂斯（Bates）等人把评价视

为一种政治活动。他们认为评价可分为技术性评价和批判性评价。前者的价值（或意识形态）以实证主义为归依，强调效率、管理和生产；后者则视学校为不同意识形态的竞争场所，评价的重点在于了解一定社会背景下学校和课程中存在的霸权（hegemony）状况，并以公义和价值作为评价准则。

贝蒂斯的观点实际上属于课程社会学或政治学的范畴。课程社会学认为，学校课程实际上是社会各种影响势力交互作用的产物。课程一般倾向于去复制社会上的生产关系，着力反映中产阶级或主导阶级的意识形态。另外，课程也可能成为抵制社会某种势力影响的工具。贝蒂斯主张对课程进行政治学的评价分析为我们提供了评价课程的新视角。

根据上文对五个较有代表性的课程评价定义的分析，我们也许可以得出如下结论：（1）课程评价的方法多种多样，教育测验或测量只是评价的重要方法之一，更不等同于评价。（2）从功能看，有形成性评价和总结性评价之分。泰勒评价观主要强调了评价的总结性功能，克龙巴赫、斯塔弗尔比姆等则侧重于形成性功能。（3）与功能相对应，课程评价的范围可以是局限于学习结果的评价，也可以对课程实施的全过程进行评价。（4）课程评价具备事实描述和价值判断两种属性，其中前者是评价的基础，后者是评价的本质，两者不可偏废。（5）既然课程是各种社会影响力交互作用的产物，所以我们能够对课程作社会学的分析。

第二节 评价的功能

当前课程评价存在的主要问题是对评价功能的认识混

乱，或者在客观上不承认评价可以促进综合功能的实现，或者误以为通过水平性评价和选拔性评价就已经可以实现评价的综合功能。在这种认识的误导下，课程评价往往出现以下偏向。这些偏向的直接后果是在学校中滥用评价的鉴定水平和淘汰选拔功能而使评价产生高利害性的副作用。

(1) 评价只重视甄别与选拔的功能，忽视评价的综合功能。

(2) 评价过分关注活动结果，忽视被评价者的进步状况和努力程度，忽视过程的评价。

(3) 评价主体单一，忽视评价主体的多源、多向的价值，尤其是忽视自我评价的价值。

(4) 评价标准机械、单一，过于强调共性和一般，忽视个性的发展和差异性。

(5) 评价内容片面，过于侧重书面考试，忽视书面考试以外的内容。

(6) 评价方法单调，过于注重量化评价和书面测验，对质性评价方法不够重视。

(7) 忽视评价结果的反馈与认同，使评价的反思、调控功能得不到充分的发挥。

(8) 评价对象基本处于被动地位，自尊心、自信心得不到很好的保护。

课程评价的功能不只是鉴定水平和淘汰选拔，也应该具有导向性的综合功能。甄别、选拔和导向是评价功能的三方面表现。三方面的整合是课程评价功能的整体，能够对教学活动的瞬时和历时、纵向和横向作出全方位的评估，真正体现课程评价的意思。但是，它们各自的功能属性又各不

相同。

1. 甄别功能

甄别功能就是依据课程的目标,通过水平测试形式,对教学质量、教师水平、学生成绩等方面进行鉴定。它是评价的最基本功能,其他评价功能都是以甄别功能为基础和前提的。如果对评价对象不先作出甄别与鉴定,评价的激励、促进等综合功能便失去了根据。而要正确地发挥评价的激励作用与促进作用,必须先通过甄别鉴定,弄清楚评价对象的优缺点、长短处,弄准确评价对象进步多大或退步多大。同样,甄别的结果也是选拔的依据,高考如此,中考也是如此。

因此,在教学活动中,我们必须重视并正确利用评价的甄别功能。有人把“过分强调甄别功能”列为亟待改革的现行课程评价方面的弊端,是不科学的,缺乏依据的。“课程评价过分强调甄别功能”这一提法反映出判断者对评价的甄别功能与其他功能之间的关系似乎缺乏应有的正确认识。课程评价无论是对学生的评价、对教师的评价还是对课程自身和学校教学工作的评价,都必须首先发挥其甄别功能。否则,整个评价会是空洞的,不符合对象实际的。这种空洞的或虚假的评价,不仅不能起任何积极的激励和促进作用,相反会使评价对象对评价的虚假结论产生反感,甚至对评价者产生厌烦之情。现在有人主张可以对学生的缺点忽略不计,坚持“无批判语的刚性策略”,对学生只能鼓励、夸奖和表扬,一味哄着学生。这种形式上的过程评价和档案评价不能解决实际问题,并且会耗费大量时间和精力,引起广大师生的厌倦。

现行课程评价的主要问题并非过分强调了甄别功能，而是由于评价往往只停留在甄别层次上，没有以甄别为基础进一步去实现其激励和促进作用。例如，评价学生甄别出了缺点，往往没有从促进学生发展的目的出发，帮助他分析认识缺点或依靠和发扬其优点，调动其积极性去克服这些缺点，甚至在甄别发现学生的缺点后，不仅不去分析产生缺点的原因，找出克服缺点的办法，反而对学生挖苦嘲笑和无情打击，使评价应有的激励与促进作用丧失殆尽。

2. 选拔功能

如果说以教学目标为标准评价学习者学业水平，体现了评价的甄别功能，那么在一个团体中通过各个个体间的横向比较，评价出该团体中的各个等次，这就是评价的选拔（或筛选）功能。高一级学校招生需要选拔，学科的竞赛也是依据评价的选拔功能，社会上的各种招聘考试大多属于选拔性的评价。

同评价的甄别功能一样，评价的选拔功能具有不可替代的作用。例如，我国一年一度的高考必须依靠评价选拔功能，按照每年高校招生人数的计划和该年中学毕业生人数的情况设定一定的分数线，选拔一批学生（尽管按照评价的甄别性，有可能有一部分考生的学业水平不高），或淘汰一批考生（有可能有一部分考生的学业水平已经达到标准）。如果我们完全撇开评价的选拔功能，借助水平考试甄别或其他综合功能来招收高校新生，那么，在高等教育也未普及之前，其结果肯定是不科学、不合理、不公平的，甚至会引发社会动荡。高考如此，中考也是如此。

但是，有人简单地把应试教育归根于各级各类的选拔考