



外语课堂评价 理论与实践

傅春晖◎著

FOREIGN LANGUAGE CLASSROOM EVALUATION
THEORY AND PRACTICE

湘潭大学出版社



外语课堂评价 理论与实践

傅春晖◎著

湘潭大学出版社

图书在版编目(CIP)数据

外语课堂评价理论与实践 / 傅春晖著. — 湘潭:
湘潭大学出版社, 2010.9

ISBN 978-7-81128-235-1

I. ①外… II. ①傅… III. ①外语教学:课堂教学—
教学评议 IV. ①H09

中国版本图书馆 CIP 数据核字 (2010) 第 177010 号

外语课堂评价理论与实践

傅春晖 著

责任编辑:王亚兰

封面设计:罗志义

出版发行:湘潭大学出版社

社 址:湖南省湘潭市 湘潭大学出版大楼

电话(传真):0731-58298966 邮编:411105

网 址:<http://xtup.xtu.edu.cn>

印 刷:国防科技大学印刷厂

经 销:湖南省新华书店

开 本:880×1230 1/32

印 张:7.75

字 数:195千字

版 次:2010年9月第1版 2010年9月第1次印刷

书 号:ISBN 978-7-81128-235-1

定 价:25.00元

(版权所有 严禁翻印)

前 言

笔者是一位一线英语教师，在教学实践中，深感目前外语课堂评价存在的问题之多及问题的重要性。目前我国外语课堂评价的主要方式还是以终结性评价为主，但这种评价方式所传递的信息往往偏离了外语课堂教学的一些事实，在很大程度上引导学生只注重对支离破碎的语言知识的记忆和理解，忽视了语言最根本的交际能力的培养和提高。对教师而言，也遏制了教师进行教学改革能动性。现代教育理论表明，教、学、评是三位一体的。外语课堂评价方式不改变，教学改革不可能得到发展。2008年，笔者和彭金定教授共同申请的项目“外语教学环境下的教师决策研究”得到了湖南省社科基金的立项支持（08YBB035）。在研究的过程中，我们发现教师决策是在周密可靠地收集信息、准确清楚地解释信息的基础上做出的，其根本目的是促进教师的教与学生的学，而这也是外语课堂评价的主要目的。之后在对湖南省教育厅“教师发展”和“生态课堂”的教改课题研究中，同样发现课堂评价于教师和学习的重要性。这些是引发笔者研究外语课堂评价的主要原因，所以本书也是“外语教学环境下的教师决策研究”的研究成果之一。

在传统的课堂教学中，“课程—教学—评价”是一种线性关系。一谈到评价，很多人马上就会想到由专门人员实施的测试。这种对评价的理解是不全面的。现代的课堂教学中，“课程—教学—评价”是一种动态循环的关系。评价和教学是一个有机的结合体。只有将课堂评价贯穿于教师的教与学生的学，才有可能

使评价积极引导教学，一方面有利于教师反思自己的教学，另一方面引导学生进行反思性学习，培养学生的元认知学习能力。

评价作为教学过程中不可或缺的重要环节，经历了测量、描述、判断和意义建构等几个不同的发展阶段，这种意义建构阶段的评价引发了教育评价从理论到实践，从模式到技术等各个层面的研究。从国际视野看，教育评价的改革已经成为国际教育改革的主题。外语课堂评价是一个由多学科理论渗透形成的跨学科的新兴研究领域，教育科学、心理科学、认知科学、社会学、脑科学、语言学与教学的相关理论为外语课堂评价的发展提供了新的理论和科学基础。它由单纯的语言测试逐步发展，并根据语言学习的特点逐步树立起多元评价观。

语言教学的根本目的是培养学生的语言能力。那么，语言教学究竟应该使学生具备一种怎样的能力？评价是基于什么样的理论？用什么样的方法和手段进行评价？什么时候评价？这些问题都是语言教学工作者感兴趣的，也是本书力图阐述的理论和实践问题。

本书由七章构成，分别从理论和实践的角度讨论了外语课堂评价。第一章是对教育评价的概述，区分了评价、测试、评估等基本概念，阐述了教育评价的发展历程、种类与模式，讨论了教育评价与教育改革的现实之争；第二章为外语课堂评价概述，从外语课堂评价的概念、功能、原则等方面进行了论述，向读者介绍了国内外外语课堂评价的研究情况；第三章为外语课堂评价的理论基础，之所以另辟一章是因为在教学实践中，笔者发现相当一部分教师对语言观、学习理论、第二语言教学理论不甚了解，教师缺乏理论知识，进行外语课堂评价只能是依葫芦画瓢，无法进行灵活的运用；第四章介绍了外语课堂评价的一般过程，特别介绍了测试类评价信息和非测试类评价信息的收集；第五章是在基于中国外语学习特点和研究者们已有的评价模式的基础上构建

出的外语课堂评价模式和操作流程；第六章为外语课堂评价实验研究，讨论了新的课堂评价模式和操作流程的有效性；第七章讨论了外语课堂评价未来的发展和挑战，指出在社会“问责”压力下，外语课堂评价的发展依然任重道远。

外语课堂评价在语言教学领域占有重要地位，近 20 年来国外在这方面的研究已取得了较大的成就，国内的研究还相当缺乏。虽然在这方面笔者进行了一些研究和探索，但因为研究水平等各方面的原因，研究成果还不太成熟，希望得到专家和同行的指教，以促进外语课堂评价的不断发展。

傅春晖

2010 年 8 月于长沙

目 录

第一章 教育评价概述

第一节 教育评价的概念	1
第二节 教育评价的发展	5
第三节 教育评价的种类与模式	12
第四节 教育评价与教育改革	31

第二章 外语课堂评价概述

第一节 外语课堂评价的内涵	42
第二节 外语课堂评价的主要功能	51
第三节 外语课堂评价的原则	56
第四节 外语课堂评价研究	62

第三章 外语课堂评价的理论基础

第一节 教育评价与外语课堂评价	80
第二节 语言观的发展	82
第三节 应用语言学理论的发展	87
第四节 学习理论的变化	95

第四章 外语课堂评价的过程

第一节 外语课堂评价的一般过程	110
第二节 外语测试评价的信息收集	131
第三节 外语非测试评价的信息收集	147

第五章 外语课堂评价模式的建构

第一节 中国外语学习的特点	164
---------------	-----

第二节	外语课堂评价模式的建构理论·····	168
第三节	外语课堂评价模式流程的描述·····	169
第六章	外语课堂评价实验研究	
第一节	研究设计·····	184
第二节	数据结果统计与分析·····	192
第三节	结果分析与讨论·····	201
第七章	外语课堂评价的发展趋势与挑战	
第一节	外语课堂评价的发展趋势·····	209
第二节	外语课堂评价的挑战·····	213
附 录	·····	217
参考书目	·····	228

第一章 教育评价概述

教育评价是教学过程中不可或缺的重要环节,是从教育测量活动中发展出来的一种价值判断活动。教育评价用来判定学生在一些教师关注的教育变量(如知识、技能、情感、态度)上的状况(Popham,1993)。教育评价作为一门学科,和其他的学科一样,也经历了一个从产生到发展的过程,是伴随着学校教育的产生、教育研究的发展而逐步发展起来的。教育评价起源于古代的传统考试,经历了测量时期、描述时期、判断时期和建构时期等几个不同的发展时期。教育评价有哪些种类?教育评价有哪些具体模式?教育评价有什么功能?教育评价的现状如何?本章将进行初步的讨论。

第一节 教育评价的概念

在讨论评价之前,我们需要明确一些基本的概念,弄清它们的基本关系,澄清一些模糊的认识。需要指出的是,在教育评价领域,评价、评估、测量、考试等基本术语的定义也是有分歧的,人们对基本术语概念的认识也不尽一致,作为教育工作者非常有必要了解大家的共识。

一、评价与教育评价

评价就是一种价值判断,是一种特殊的认知活动(鲁子问、王笃琴,2005)。评价就是评定价值的高低,即给予某事物或人以价

值上的判断。评价也解释为行为方式的描述加上行为方式的价值判定(冯平,1995)。这些定义都跟价值有着密切的关系。要了解什么是评价,首先要了解什么是价值。价值是人类对于自我发展的本质发现、创造与创新的要素本体,包括任意的物质形态。有人认为事物本身是没有价值的,价值是主体需要的表达。价值并不依赖于评价而存在,评价只是揭示价值关系,并不创造价值。

价值在很多领域有特定的形态,如社会价值、个人价值、经济价值、法律价值等。这些价值的存在是人在不同领域发展中范畴性、规律性本质的存在。要了解什么是教育评价,首先要了解什么是教育价值。美国学者 N. E. 格朗兰德(N. E. Gronlund)在 1971 年以如下公式表达了教育评价的概念:

评价 = 测量(量的记述)或非测量(质的记述) + 价值的判断

教育活动满足社会群体或个体的需要就是教育价值,而教育价值有教育的社会价值与教育的个体价值之分。“价值”这一概念与“需要”有着密切的联系。社会对教育的需要是一种特殊的需要。社会对教育的需要由个体(受教育者及其家长)对教育及国家或地区对教育的需要两部分组成。

广义的教育评价是指用系统科学的方法全面收集信息、分析信息,从而对教育事实作出价值判断的过程。狭义的教育评价是指依据教育目标对学生进行教育后产生的行为变化作出判断,并为改善和优化教育途径提供依据的过程(王孝玲,1999)。教育评价是一个教育科学的概念,最早由美国教育学家 R. W. 泰勒(R. W. Tyler)正式提出,它在教育领域中是一个宽广的概念,可涵盖这一领域中的所有测试、测量和评估行为,是人们价值活动不可缺少的一部分。在我国,教育评价研究起步较晚,自 20 世纪 80 年代开始,广大教育研究者和教师加强了对教育评价的研究,教育评价得到了迅速发展。但到目前为止,教育评价还没有一个确切的、严谨的、被一致接受的科学定义,比较认可的说法有以下几种:

(1)教育评价就是以教育为对象,对其效用给予价值上的判断。

(2)教育评价是利用所有可行的评价技术评价教育所预期的一切效果。

(3)教育评价是对照教育目标,对由教育行为产生的变化进行的价值判断。

(4)教育评价是人们按照一定社会的教育性质、教育方针和教育政策所确立的教育目标,对所实施的各种教育活动的效果、学生学习质量及发展水平进行科学的判定。

从以上的定义可以看出,教育评价的关键还是在于教育活动效果的价值判断。但对于多数非教育工作者,甚至是少数教师而言,评价、评估即等于测试。人们往往认为评价、评估和测试是可以互换的术语。王蔷(2001)指出,评价常与测试相联系,每当谈到评价的方法时,许多教师马上就想到测试。然而,评价与测试之间有一点是很不相同的,评价通常是根据搜集到有关学习者当前情况的信息做出的,而测试只是搜集信息的方法之一。

二、测试、评价与评估

“Test”一词可翻译为测试或测验。“测试”这一术语在学术著作中使用得很普遍,而“测验”这一术语在学校的教师和学生中使用得较普遍,本书使用“测试”这一术语。在教育评价领域,它们是人们最为熟悉的测量工具,能引起人的代表性行为,以便对人的行为特性或心理特性进行测量和评价。有学校教育经历的人,都从老师手中接到过一份又一份的试卷,这些测试的目的是想确定学生们对于老师所教知识和技能的掌握水平。测试作为评价工具的一种,经常被用来收集评价信息,但测试所收集到的信息相对而言也是有限的,它只能告诉我们学习者成绩的某一方面。测试与评估、评价不同,它可以为评估和评价提供数据。

在《双解英汉字典》中对“评价”的解释是:估计、求出……的数量或价值。汉语“评价”是评定价值的简称。在英语中有评价之义且容易混淆的两个表达分别是“Assessment”和“Evaluation”。

评估(Assessment),亦称评定,是一个系统的、多步骤的过程,涉及信息的收集和解释,旨在了解学习者知道什么以及能做什么(Tamir,1998)。在教育领域中,评估包括通过观察、测试、面谈、问卷以及其他测量手段所获取的定性、定量信息并由此来判断学习者在某一学习过程中的成绩和学习表现(Doran, Lawarenz & Helgeson,1994; Mitchell,1992)。P. J. 希尔斯(P. J. Hills)在编写的《教育词典》(*Dictionary of Education*)中写道:评估主要的重点是评定学生学习的进展。评估趋向于采用质性评价的方法,以及定量、定性评价相结合。其狭义的理解是指不涉及考试在内的评估程序(Davies,1999/2002)。大多数的研究者(Little & Wolf,1996; Brown & Hudson,1998; Gipps,1994; Brindley,2001)认为,评估具有更广泛的意义,它包括测量、测试、评价等活动。G. 布林德利(G. Brindley,2001)认为“评估”一词是没有任何限制的;吉普斯(Gipps,1994)认为评估包括测试、考试、课程作业、实践、教师观察等。

“Evaluation”,在我国教育理论界通常译作“评价”,它所关注的重点是学生是否正在学习他们被期待学习的东西。胡森等主编的《国际教育大百科全书》建议保留这个术语,使其专门用于抽象的存在物,比如方案、课程以及组织变量中。在我国正式出版的学术著作中作使用“评价”这一术语的占多数,而教育主管部门颁发的文件则较多地使用“评估”这一术语。

对于评价和评估的区别,学者们看法不一,有不少学者认为这两个术语可以交替使用(Montgomery,2004),因为他们认为评价与评估是一个交互进行的过程,本书采用“评价”这一术语。

第二节 教育评价的发展

一、西方教育评价的发展历程

早在两千多年前我国战国时期就出现了教育评价的萌芽,但教育评价制度和理论的真正形成和发展却是在 19 世纪后半期的欧洲。教育测量序幕的拉开始于 1869 年,英国的 F. 高尔顿(F. Galton)通过对个体差异的长期研究,出版了《遗传的天才》一书,提出了天才由遗传决定的观点。在此之后,国外一些实验心理学家、数学家等摸索出了一些测试和测量的方法(Wandt, 1879; Oehro, 1889; Pearson, 1896)。1904 年,美国的斯泰克(Stake)出版了《心理与社会测量导论》一书,介绍了统计方法和编制测试的基本原理。斯泰克提出了凡存在的东西都有数量,凡有数量的东西都可以测量的基本观点。因此,人们认为通过“测量”可以检验教师教学成效的大小、学生对知识记忆的情况。所以,这个时期的评价就是选择测量工具、组织和实施测量、提供测量数据。因此,这一时期被称为测量时期,为后来教育评价成为一门学科奠定了研究基础,但这并不属于真正意义上的评价。

20 世纪 30 年代以前,现代教育评价的概念尚未形成,教育测量与教育评价几乎是同义语。美国教育学家泰勒首次提出了教育评价的科学概念,使得教育评价作为一项科学来研究。为了促进和保证当时美国课程改革的进行,美国进步主义教育协会进行了一项课程内容改革的实验研究,从 1932 年到 1940 年历经 8 年时间完成,史称“八年研究”。泰勒认为:评价不仅仅是一两个测试,而是根据预定教育目标对教育结果进行客观描述的过程。他建议用“评价(Evaluation)”一词取代“考试(Examining)”和“测试(Testing)”。以目标为中心,通过具体的行为变化来判断教学目标实

现的程度是泰勒教学评价的实践方式。试图通过多种手段,描述学生的学习结果与教育目标的一致程度,从而发现问题,改进课程教材、教育教学方案和方法(李亦菲,2004)。同时,泰勒还提出评价的原则和方法,即“泰勒模式”。“八年研究”实际上宣告测量运动的终结,人们在思想上和行动上接受了采用教育评价方法来描述教育成效这一观点,将评价理解为对目标实现程度的描述,形成了一个以“描述”为特征的评价时代。

到了20世纪50年代,1957年美国因为前苏联第一颗人造卫星上天而大为震惊,进而进行了大规模的教育改革,教育评价也受到新的质疑和挑战(李亦菲,2004)。人们对“泰勒模式”开始置疑,主要关心以下问题:

(1)对已经确定的目标,是否需要评价?是否需要价值判断?

(2)判断是否应成为评价的一项基本活动?

(3)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?

1963年,克龙巴赫(Cronbach)发表了一篇题为《通过评价改进课程》的文章。1967年斯克里芬(Scriven)发表了论文《评价方法论》,同年,斯泰克发表了《评价的面貌》一文,肯定了判断是评价的两大基本活动之一,同时提出了一个完整的、包含描述与判断两个方面的评价模式。这一阶段确认了价值判断是评价的本质,确认了评价的过程性,强调评价应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。这一时期将评价理解为价值判断,所以这一时期被称为判断时期。

20世纪70年代初,教育研究人员在对美国20世纪60年代课程改革运动深刻反省中,意识到前几个阶段的教育评价忽视了价值的多元性和学习者在评价中的主体地位,人们渴望发展新的评价理论和范式。

20世纪80年代,美国的E. 古巴(E. Guba)和Y. S. 林肯

(Y. S. Lincoln)联合发表了《有效的评价》(1981)和《自然主义的研究》(1985)两篇论文,1989年出版了名为《第四代教育评价》的专著。在这本书中,他们认为评价是评价者与评价对象双方通过平等交互作用共同建构统一观点的过程,评价结果也是双方交互作用的“产物”。他们提出在评价中存在“价值差异”,认为参与评价的人的价值观是各不相同的、价值标准是存在差异的。根据这种理解,他们认为评价是一种民主协商、主体参与的过程,应在自然的条件下,使各方人士通过各种形式的对话达成共识,而不是评价者对被评价者进行控制的过程。这一阶段的一个重要特点是强调质性方法的应用,力图通过自然的调查,全面充分地揭示和描述评价对象的各种特质,将评价理解为意义建构,这一时期被称为评价的建构时期。

《第四代教育评价》在全世界教育领域中产生了很大的反响。“将评价理解为意义建构”这一观点已经成为教育评价改革的基本思路。以意义建构的评价理念为基本支撑点,灵活采用多样化的评价方法开展评价,力图建立促进学生发展的评价。这一时期产生了真实性评价、行为表现评价、成长档案评价等新的评价理念。

二、中国教育评价的发展历程^①

自古以来中国就是一个具有丰富教育思想的文明古国。中国教育评价活动在古代一直与选士制度密切关联。西周的选士制度是世界上最早的通过评价选拔人才的制度,在春秋战国时期,有了“学而优则仕”的说法。西汉建立以察举制为核心的选士制度,标志着中国古代选士制度进入了一个新的时代。教育评价的重要方式——考试,在察举制中正式诞生。自隋朝开始的科举取士、分科

^① 这里主要参考了刘尧(2003)和高凌飏(2009)的研究成果。

选拔、逐级考试,在考生来源、考试科目与方法、录用程序等方面都形成了一套较完备的制度,这也可以看作有系统的教育评价活动的开始。一些西方教育评价的研究者(Dore, 1976; Sposky, 1999)认为,中国是世界上第一个通过学业考试来选拔政府官员的国家,这被认为是社会进步的一大标志。科举制的问世,开创了人才教育评价、选拔的新纪元。明代科举制的高度程式化(八股取士)及古代“标准化”考试的问世,使科举制度得到进一步的强化,笔试是选拔政府官员的唯一固定的方式,八股文成为通往仕途的必经之路。中国的教育评价萌芽较早,但古代科举制度的高利害性、甄别选拔性,束缚了评价制度朝科学性、系统性的方向发展。但不可否认,中国古代的考试制度以其鲜明的特点、强大的生命力和作用力对中国乃至全人类社会的发展产生了极为深远的影响,它对今天的教育评价仍有积极的借鉴意义,这是现代教育评价本土化的生长点。其借鉴意义表现在如下方面:

- (1)教育评价是追求教育公平的主要手段;
- (2)教育评价的管理方法要追求科学化;
- (3)教育评价的方式要追求多样化;
- (4)教育评价要与学校教育活动配套衔接。

古代中国所萌发的教育评价思想的嫩芽,却因种种原因未能得到科学、系统的研究,自19世纪以来,一些西方国家陆续建立了文官考试制度,与我国的科举考试基本相似。20世纪,西方教育测量和评价运动率先成长发展起来。1905年,中国废止了承袭已达1300年之久的封建科举制度之时,正值西方教育测量运动方兴未艾,在内与外、主动与被动两种力量的共同作用下,西方教育测量的理论很快就传入了中国。20世纪一二十年代,一些激进的学者开始引进西方的心理测试方法,这期间,以《教育与心理测试》和《教育测试》为主要代表,进而在20世纪二三十年代形成了中国的教育测量运动。伟大的“五四”运动,使国民思想空前解放,

西方教育思想大量涌入。中国早期的教育理论工作者翻译引进国外的测验量表,1931年成立了中国测量学会,开展测验理论研究,全国各大学教育系和中等师范学校普遍开设了教育测量学的课程,理论研究取得了进步。中国测量学会成立后,关于测量的理论研究得到了加强,一些有关理论探讨的文章得到发表。教育评价与教育测量相结合,并呈现出多种体制的多元化格局。新中国成立后,测量理论从社会教育、政治教育和学校教育三个方面提出了明确的教育目标和要求,成为发展教育的方向和评价教育的准则。但20世纪70年代的“文化大革命”中断了正常的中国教育,这个时期的教育评价研究没有任何进展。

从20世纪80年代初开始,中国的许多教育期刊陆续译介了国外及中国台湾地区有关教育评价的文章及专著,邀请外国教育评价专家来中国讲学。这些活动全面介绍了世界教育的动向、发展趋势以及国际教育评价研究与实践活动的动态,加强了中国同国外教育评价界的联系和交流。随着教育改革的深入,在系统引进和学习国外教育评价理论和方法的基础上,教育评价实践活动在全国有组织地展开。1984年1月中国正式签署了入会文件,加入“国际教育成就评价协会”。教育部指定在中央教育科学研究所建立“中国国际教育成就评价中心”,教育评价工作有组织地展开,教育评价实践的种类也逐渐增多,从较为单一的教学质量评价,发展到对学术水平、学科发展方向、学校后勤工作、实验室管理和学校校办工厂等的评价。1985年5月,中共中央颁布了《中共中央关于教育体制改革的决定》,明确提出要对教育进行评价的问题。于是,在全国开展了教育评价研究和试点工作,探索评价规律,建立评价理论和方法体系,为评价工作的全面展开铺路。在原国家教委的统一领导和指导下,在全国范围内全方位、多层次地开展了各种类型的教育评价实验和实践活动。教育评价研究的内外交流活动得到进一步加强,并且在学术交流方面也产生了质