

语言学及应用语言学系列丛书



教育部人文社会科学重点研究基地



广东外语外贸大学外国语言学及应用语言学研究中心

全国基础教育阶段英语科 监测系统的后效研究

亓鲁霞 著

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

语言学及应用语言学系列丛书



教育部人文社会科学重点研究基地



广东外语外贸大学外国语言学及应用语言学研究

全国基础教育阶段英语科 监测系统的后效研究

亓鲁霞 著

外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

北京 BEIJING

图书在版编目(CIP)数据

全国基础教育阶段英语科监测系统的后效研究 / 亓鲁霞著. — 北京: 外语教学与研究出版社, 2018.6

(语言学及应用语言学系列丛书)

ISBN 978-7-5213-0116-8

I. ①全… II. ①亓… III. ①英语课—教学质量—监测系统—研究—中小学
IV. ①G633.412

中国版本图书馆 CIP 数据核字 (2018) 第 135795 号

出 版 人 徐建忠
责任编辑 赵东岳
责任校对 孔乃卓
封面设计 覃一彪 高 蕾
出版发行 外语教学与研究出版社
社 址 北京市西三环北路 19 号 (100089)
网 址 <http://www.fltrp.com>
印 刷 北京九州迅驰传媒文化有限公司
开 本 650×980 1/16
印 张 10.75
版 次 2018 年 6 月第 1 版 2018 年 6 月第 1 次印刷
书 号 ISBN 978-7-5213-0116-8
定 价 42.90 元

购书咨询: (010) 88819926 电子邮箱: club@fltrp.com
外研书店: <https://waiyants.tmall.com>
凡印刷、装订质量问题, 请联系我社印制部
联系电话: (010) 61207896 电子邮箱: zhijian@fltrp.com
凡侵权、盗版书籍线索, 请联系我社法律事务部
举报电话: (010) 88817519 电子邮箱: banquan@fltrp.com
法律顾问: 立方律师事务所 刘旭东律师
中咨律师事务所 殷 斌律师
物料号: 301160001

本书的研究得到国家社会科学基金一般项目“全国基础教育阶段英语学习质量监测系统研究”(10BYY031)的资助。

课题组成员包括：亓鲁霞，李子容，李金辉，武尊民，黄志红，蔡宏文，张宪，谢琴；参加研究的人员还有徐柳，张莉，袁丽娟，陈小聪，彭震，许艺，晏盛兰等。

编委会

顾问：桂诗春

主任：刘建达 董燕萍

编委会成员（按汉语拼音排序）：

陈建平	董燕萍	杜金榜	何晓炜	黄国文
李绍山	李行德	刘建达	亓鲁霞	冉永平
沈家煊	束定芳	王初明	温宾利	文秋芳
吴旭东	徐海	曾用强	章宜华	郑立华

总序

“语言学及应用语言学系列丛书”是广东外语外贸大学外国语言学及应用语言研究中心（教育部人文社科重点研究基地）策划出版的一套丛书。

改革开放以来，我国迎来了科学的春天。1981年国务院学位委员会把语言学和应用语言学列入学科目录，并开始逐步在我国建立硕士和博士学位授予单位。这是学科发展的需要，也是历史的必然。从《圣经》“通天塔”故事到今天网络的蓬勃发展，都说明语言和语言交往（包括语言的习得和教学）是人类社会不可或缺的一个极为重要的元素，也预示着语言学和应用语言学的强大生命力。语言和人类生活休戚相关，往往被称为人们身边的科学。每一个人都有学习和使用语言的经验 and 体会：就怎样使用语言而言，凡是受过教育的人大都可以判断哪些话可以或不可以这样或那样说；就语言习得而言，每一个人都经历过大致相同的过程，最后都学会了说话。至于怎样学另一种语言，大多数人都有一些自己怎样学习的门道，有自己的话语权，可惜的是学习的结果似乎没有母语学习那么简单和显而易见，而是有成功有失败。这里面有些什么玄机呢？Dave Willis（2003: 1-6）在他的近著里指出：作为教师，我们看到学习是以一种难以预测的方式进行的，我们“**所教的东西不一定是[学生所]学到的东西**”（What is taught may not be what is learnt）。而学生往往又会学到一些不是教师所教的语言现象。他曾做过一个小试验，把一个在职的师资培训班里的教师分为两组，让他们分别列出一年级学生和三年级学生最常犯的10种英语失误。事后他把失误加以比较，发现有7种竟然是完全一样的（例如不会使用冠词、第三人称动词的现在时态，不会使用do助动词来提问，等等）。这自然引起一些关于课堂教学的严重问题：教师是否真的要用两年时间才能减少三种失误？是否三年级学生并不比一年级学生好多少？我们怎样解释这种可怕的失败？Willis所提出的这些问题发人深思，一个简单的回答是：学习者在学外语的时候面临着和母语习得很不相同的问题，包括语言学习观、学习动机、学习态度、学习方法、学习策略、注意、智力、情感、学能……等等因素。学习者的这些差异性在不同时候会在学习者的个体身上产生不同的作用。而另

外一方面,我们作为实施教学的一方(从教育部门的领导者、组织者到教师等)往往以为“所有的受教育者都是一样的”(Skehan, 1998: 260-261),因而忽略了他们的差异性,只顾制定和使用统一的教学大纲、教材,提出统一的要求;连教师的培养都是为了“巩固这些隐含的权力关系”。

这个例子说明,应用语言学已经发展成为一门跨学科的科学,必须从语言学、心理语言学、社会语言学、神经语言学、语用学、语篇分析、教育学等角度去综合解决语言教学中所出现的种种问题。现再从语言学中的语言起源这一个特殊问题来说明其多学科性。大家都知道,巴黎语言学会曾于1886年明令拒绝接受关于语言起源的论文,因为虽然这个问题曾经不断激发许多代人的想象力、兴趣和注意,但是所有的研究都陷于无休止的、无结果的思辨性讨论。可是到了20世纪50年代以后,由于语言学家、认知科学家、脑神经学家、心理学家、分子生物学家、进化生物学家、解剖学家、人类学家、考古学家的共同协作,这个问题的研究取得了很大进展,大大地开拓了人们的视野。哈佛大学心理学家Hauser、Fitch和麻省理工学院语言哲学家Chomsky(2002)在*Science*所发表的他们的语言机能新观(语言机能有关义和狭义之分,而狭义语言机能则只有递归)就是基于这些研究。在文章里,他们提出语言进化的比较方法——根据现存物种的实证性数据来详细地推断灭绝的先祖。这篇文章引起了一场热烈的讨论。Pinker和Jackendoff(2005)是他们的对立面,也提出很多甚有见地的看法。他们并不反对语言机能有关义和狭义之分,而且对Hauser和Fitch关于别的物种语言的相似性(analogs)和同源性(homologs)的研究表示赞赏。但是他们坚持主张语言是人类在进化过程中自然选择的结果,是为了适应交际的需要而产生的。他们也同样是在多学科的背景下展开讨论。并非针对Hauser等人的文章,而单独从不同的角度对语言和交际起源而展开富有成果的研究的,还有哲学家Dan Dennett(1995)、生物学家Richard Dawkins(1976)、Peter McNeilage(2008)、人类学家Gordon Hewes(1973)、进化生物学家Robin Dunbar(1997)、发展心理学家Michael Tomasello(2008)等。

正是基于这种秉承“质量高、彰显‘新’”(新议题、新视角、新观点、新方法等)的理念,广东外语外贸大学外国语言学及应用语言学研究组织了这套“语言学及应用语言学系列丛书”,第一个系列有4本著作,它们是《英汉双及物结构的生成语法研究》(何晓炜)、《计算机化考试研究》(曾用强)、《英语学习型词典研究》(徐海)、《词汇语用探新》

(冉永平)。这几本新著篇幅都不是很长(十多万字),但却要言不烦,有血有肉;少的是泛泛之谈、言不及义,多的是个人的实际观察和亲自试验。而且选题得当,又经过爬梳剔抉,信而有征,对推动我国语言学研究,起了添砖加瓦的作用。该系列丛书是一个开放的系统。希望第一系列投放后,更多的作者能够关注并踊跃支持该套丛书,使我国语言学研究能够绽放出更多的奇花异草,春色满园。

桂诗春
2011年8月

序

这是一本关于测评后效研究的书。所谓后效，常指测评对教学、教育系统乃至社会造成的影响。一些高风险考试（如高考）影响巨大，对教学既可能产生积极的推动作用，也可能带来消极的误导，因此，后效研究受到测评界高度关注并成为测评的一个研究分支，实属顺理成章。然而，本书研究的对象并非高风险考试，而是全国基础教育阶段英语科教学质量监测系统，旨在跟踪调查质量监测的潜在影响，而非已形成的实际后效。

全国基础教育阶段教学质量监测于2007年正式启动。在教育部的直接领导下，依托于北京师范大学，我国成立了基础教育阶段教学质量监测中心，负责全国监测工作的实施，英语教学是监测科目之一。提质增效是各行各业的生存之道和永恒追求，教育自然也不例外。教育的重要功能之一，是通过课堂教学传授知识，培养技能，提高学生素养。衡量教学质量优劣的最主要指标理应看教学大纲的完成情况，最终落实到学生对大纲规定的知识和技能所掌握的程度。而影响学习效果的外部条件也十分重要，涉及师资队伍、教学理念及方法、教学水平、教学配套设施等因素。教学质量监测就是针对所有这些因素进行定期评估，掌握教学状况的发展变化，以利于教育的科学决策，规避拍脑袋瞎指挥。

英语科是质量监测的主要内容之一，而决定开展后效研究，主要是由监测方式决定的。监测采取了测试与问卷调查相结合的方法，借此全面了解英语教学情况以及对其产生影响的相关因素。既然监测用了测试，势必对教学具有潜在的反拨作用。虽然监测结果一时不触及学生和教师的切身利益，对教学影响不大，但是，随着监测的深入持久开展，随着监测持续反馈给各级政府部门，监测结果会否被误用于评教评学，对师生形成压力，引发应试教学，额外增加测评风险？对此可能性实难预料。为了促进监测工作的顺利进行，我们提出，一个保险的措施是未雨绸缪，在外语教学质量监测过程中对监测本身实施监测，采用的手段就是开展后效跟踪研究，这样做既检视教学质量的现实状况，又着眼于监测工作的未来发展，形成良性循环。我们的建议很快获得上级部门的同意并付诸实施。

为了将后效跟踪与质量监测有机结合起来，我们充分利用了自己后效研究的专长，成功申请到一个国家社科基金课题，并围绕课题组建了研究团队，在监测统一框架的指导下开展研究。根据后效研究的特点，课题组规划了调查进程，制定了具体实施步骤，在监测中心抽样的基础上深入中学英语教学第一线，按部就班开展一系列实地考察，包括听课、访谈、录音录像、问卷调查等，据此收集到宝贵的基准数据，并对数据进行了整理分析和结果归纳，描述监测开展之前和初期的教学现状，为今后教学质量的变化进行有效对比做好铺垫。以上过程在本书中均有详细描述，具有可复制性。

需要提及的是，书中报告的内容主要是监测之前和初期的工作，属于后效基准研究，这在以往国内外的后效研究中常见缺省，无基准的后效研究难免给结论的科学性打折扣。教育质量监测事关国家教育的健康发展，一开始便锚定基准，着眼于长远，极具前瞻性。此外，将后效研究延伸至教学质量监测，不仅拓宽了语言测试后效研究的应用范围，而且为将来的质量监测跟踪打下了坚实的基础，同时提供了可操作的方法。这样的研究对其他科目（如语文和数学）的监测具有借鉴意义，对从事测评研究的学者和研究生也有参考价值，更为那些想了解我国中学英语教学现状和质量监测情况的读者提供了较为详实的资料。因此，我们决定为读者奉上此书，以期丰富我国的测评后效研究文献。

亓鲁霞

2017年10月31日

前言

本书基于国家社会科学基金一般项目“全国基础教育阶段英语教学质量监测系统研究”(10BYY031)。该课题针对英语科基础教育质量的监测系统而开展。有两个认识推动了这项工作,其一,基础教育质量监测是一项长期的任务,将成为我国基础教育领域的常规措施。在我国开展这项工作,既适应了人才强国的战略需求,也与国际趋势吻合。当今世界,基础教育质量监测受到各国教育部门的高度重视。美国的基础教育质量监测项目NEAP已实施了六十多年,至今仍在继续,国际上的相应项目PISA也方兴未艾。我国的基础教育质量监测虽然起步较晚,但受到国家的高度重视,2007年成立了教育部基础教育质量监测中心,2012年由北京师范大学牵头,联合多所部属师范大学和教育及考试机构,成立了中国基础教育质量监测协同创新中心,定期开展教育质量监测工作,完善我国基础教育质量评价体系 and 提升我国基础教育质量。其二,大规模测评容易产生后效,特别是负面后效,如应试教育、考试作弊等,这已被大量实证研究所证实。从发展的眼光看,测评工作是动态演变的,涉及诸多因素,它们交互作用,最终变化结果难以预测。为此,本课题赶在英语科监测实施之前启动,旨在探讨英语科质量监测系统对中小学英语教学的潜在及后期影响,为监测系统提供反馈信息,使其不断得到改进,避免产生对教学的负面影响,从而保证监测工作顺利持久进行,达到促进教学改革、提高教学质量的目标。

课题的研究思路参照了当今国际流行的动态系统理论(Dynamic System Theory,简称DST)。该理论把事物视为复杂的动态系统,系统由子系统组成,它们之间持续地相互作用,系统内部与其所处环境的外部因素互动并置换资源,以此获得发展动力。由于系统的运作复杂,其变化不是简单的单向因果关系(如X导致Y),而是双向甚至多向的非线性因果关系。系统中某个节点的一个小变化,其效应在多因素的交互作用下有可能不断放大,导致系统剧烈变动,这种情况被称为蝴蝶效应。但是,尽管系统在其发展变化过程中呈现不确定性和不可预测性,但它具有自适应性和自组织能力,有时会在一定的时间段保持相对平衡和稳定。系统的这些特征对我们开展教学质量监测的后效研究有重要的启迪作用,

打开了研究思路。监测过程可视为一个动态系统，对它进行跟踪研究，探究它与环境的交互作用，揭示它的演变特征与发展规律，可以更好地发挥相关职能部门的管控和协调系统的功能，规避测评给教育发展带来的潜在负面影响。

课题的具体操作分两步进行。课题组首先对监测系统的设计者进行了观察并分析了相关文献，了解系统初期的风险构成和期望后效。第二步是到抽样学校进行实地调研，了解小学四年级和初中二年级的英语教学状况以及师生对监测的态度，藉此建立基准参照点，用于今后对系统的风险构成和教学的变化进行长期的跟踪研究。课题组对调研数据进行了信、效度验证和统计分析，建立了三个有利于后期跟踪研究的参照点。

第一个参照点是当下英语科监测系统的风险构成和设计者的期望后效。从风险构成的分析结果来看，英语科监测系统目前属于低风险考试，得到广大师生的支持。期望后效的调研结果显示，英语科监测系统设计者们期盼充分利用监测，促使教学大纲和课程标准得以落实，学校重视培养学生的兴趣和英语综合运用能力，推动中小学英语教学良性发展。

第二个参照点是小学四年级英语教学的现状。调查结果显示，大多数小学四年级英语教师基本上按照大纲和课标的要求开展教学，赞同采用接近真实语言使用的教学活动，也赞同培养英语综合运用能力的教学原则，但在实际教学中仍有相当一部分教师采用传统的语言知识灌输法，忽视学生的兴趣和语言运用能力的培养。小学四年级英语教学的最大问题是课时普遍不足，许多学校的教学设备落后且匮乏。

第三个参照点是初中二年级的英语教学现状。研究表明，初二英语教学距大纲和课标的要求甚远。受当今外语教学理论的影响，虽然初二英语教师赞同教学大纲提出的要求，也赞同采用接近真实语言使用的活动开展教学，但是他们中的大多数在实际教学中却无法摆脱以语言知识为中心的传统教法。尽管初二的教学时间比较充足，教学设备也不差，英语课本也较好地体现了大纲和课标的要求，然而，课本中不少有助于培养英语综合运用能力的好练习却被弃之不用，取而代之的是语言知识灌输与机械操练。这与中考英语有关，相当一部分课内外练习针对中考的内容和题型来设计，应试教学倾向比较严重。

本研究的主要贡献有以下三点：其一，上述三个参照点以及本课题开发的研究框架和工具，有助于英语科监测系统后效的跟踪调查，也能为语文数学等其他科目的监测后效研究提供借鉴与参考，这是本课题对监测工作的一个贡献。其二，本研究将动态系统理论（DST）应用于后

效研究，这在国内外尚属首次。DST不但较好地解释了一些后效现象，还确立了测评系统的风险构成及其演变的研究范式，能够深化我们对后效源头的认识，为后效研究打开了进一步发展的空间。其三，本课题在英语科监测实施之前便启动了后效基准研究，这在大规模语言考试和教学质量监测评估史上应属首例。研究测评后效，需要对比不同时间节点收集到的教学状况数据，而测评之前或之初的调查数据对建立测评与教学之间的复杂因果关系格外重要。综观国内外的后效研究，能够收集到测评系统建立之前基准数据的尚未见有报道，足以突显本课题的学术和应用价值。

本书的第一章阐述了课题的研究目的及意义。第二章是文献综述。在回顾梳理了有关后效的研究之后，我们认为后效是一个异常复杂的社会现象，涉及的因素庞杂繁多，诸因素之间的关系错综复杂，难以理顺，以往的研究尚欠深入，需要寻求新的突破方向。鉴于此，有必要尝试新的理论探索，推进后效研究。于是我们介绍了国际上应用语言学界当下极为推崇的动态系统理论，借鉴该理论指导本课题的开展。第三章提出课题的研究设计与描述具体研究方法。第四至六章呈现研究结果，首先报告英语科监测系统的风险构成与期望后效，以及涉考者对监测的了解与态度，然后分别描述小学英语四年级和初中英语二年级的教学现状，作为后效跟踪研究的参照点。第七章总结了研究的主要成果与局限，并提出了后续研究框架和可用工具。

最后，作者衷心感谢课题组成员：李子容、李金辉、武尊民、黄志红、蔡宏文、张宪、谢琴，以及参与收集分析数据的研究生：徐柳、张莉、袁丽娟、陈小聪、彭震、许艺、晏盛兰等。他们为课题研究付出了宝贵的时间和辛勤的劳动。此外，还有许多人员对课题研究提供了大力支持和无私的帮助：教育部基础教育质量监测中心的有关人员，江西、广西、广东的一些英语教研员，接受调研学校的学生、教师和领导，以及上网填写调查问卷的教师和学生，在此也对他们表示衷心的感谢；同时还感谢广东外语外贸大学外国语言学及应用语言学研究对本课题的支持。

亓鲁霞
2017年11月

目 录

第一章 研究目的与背景	1
1.1 研究目的及意义	1
1.2 国外基础教育质量监测	4
1.3 我国的基础教育质量监测及英语科监测系统	5
第二章 后效研究与动态系统理论	7
2.1 后效研究综述	7
2.1.1 理论探讨	7
2.1.2 实证研究	12
2.1.3 存在的问题	18
2.2 动态系统理论与后效研究	19
2.2.1 动态系统理论 (DST) 简介	20
2.2.2 DST 的主要研究路径与方法	22
2.2.3 后效: 考试与教学及社会交互作用的结果	24
2.2.4 DST 视角下的后效研究反思	26
第三章 研究设计与方法	30
3.1 研究总设计	30
3.2 研究方法和数据收集工具	32
3.3 数据来源与收集方法	34
3.4 数据分析	36
3.4.1 参与者观察	36
3.4.2 访谈数据	36
3.4.3 课堂观察数据	37
3.4.4 问卷数据	37
第四章 考试风险与期望后效 (子课题 1)	47
4.1 考试系统的风险构成及变化	47

4.2 考试风险与后效性质	51
4.3 英语科监测系统风险的构成与期望后效	57
4.3.1 风险构成	57
4.3.2 涉考者对监测的了解与态度	58
4.3.3 期望后效	61
4.3.4 内在冲突	62
第五章 小学四年级英语教学现状 (子课题 2)	63
5.1 教师信念	63
5.2 教学活动及教材设备	67
5.2.1 各类教学活动的频率	69
5.2.2 教学材料和教具	70
5.2.3 课堂举例	74
5.3 时间安排	76
5.4 测试评估及其对教学的影响	77
5.5 讨论	81
第六章 初中二年级英语教学现状 (子课题 3)	85
6.1 教师和学生的信念	86
6.2 时间安排	89
6.3 教学活动及教材与教具	89
6.3.1 各类教学活动的频率	91
6.3.2 教学材料和教具	93
6.3.3 课堂举例	97
6.4 测试评估	100
6.4.1 中考影响下的初中二年级英语测试	100
6.4.2 中考和期末统一考试对初中二年级教学的影响	103
6.5 讨论	104
第七章 研究总结	107
7.1 研究主要成果与局限	108
7.2 后续研究框架与工具	110

参考文献	114
附录	125
附录 1 访谈提纲	125
附录 2 调查问卷	129
附录 3 问卷数据收集的指令语	145
附录 4 课堂观察记录表	147
附录 5 WinMax 赋码举例	148
附录 6 课堂观察数据分析框架表	149
附录 7 问卷 2 的修改版	150

图表目录

图 1	后效模式 (Bailey, 1996).....	10
图 2	反拨效应性质、可变性及强度	11
图 3	考试风险与期望反拨效应的关系	53
表 1	研究设计	31
表 2	参与课题人员	36
表 3	小学四年级英语教学现状问卷数据因子分析结果	38
表 4	初中二年级英语教学现状问卷数据因子分析结果	41
表 5	填写问卷1的中小学教师情况 (%).....	43
表 6	填写问卷2的小学四年级教师情况 (%).....	44
表 7	填写问卷2的初中二年级教师情况 (%).....	45
表 8	教师和学生现实生活中使用英语交流的频率 (%).....	46
表 9	不同考试的风险等级	48
表 10	教师和学生对监测的了解与态度 (%).....	60
表 11	小学阶段英语学习的长远目标与教学中的期望行动	65
表 12	课堂活动及教材和教具	67
表 13	四类教学活动因子得分均值	69
表 14	小学四年级不同类型的英语教学活动频率差异比较	70
表 15	小学四年级英语课所用教学材料	71
表 16	小学四年级英语课所用教具	73
表 17	小学四年级课堂活动与气氛 (时间比例).....	75
表 18	小学四年级英语课测试评估类型	77
表 19	小学四年级英语测试考查各种内容的频率	78
表 20	六份小学四年级英语测试内容分析	79
表 21	区和市统一抽测对教学的影响	80
表 22	初中阶段英语学习的长远目标与教学中的期望行动	88
表 23	初中二年级英语课堂活动及教材和教具	90
表 24	初中二年级英语课内课外教学活动的频率	92
表 25	初中二年级不同类型的英语教学活动频率差异比较	93
表 26	初中二年级英语课所用教学材料	93