中国外语知名学者文库

语言测试中的项目分析与等值技术：研究与应用

张　权　编著

高等教育出版社
Higher Education Press

语音驱动的嘴形分析与合成技术：研究与应用

[redacted]

[redacted]

中国外语知名学者文库

语言测试中的项目分析与等值技术：研究与应用

张　权　编著

高等教育出版社
Higher Education Press

图书在版编目(CIP)数据

语言测试中的项目分析与等值技术:研究与应用/张权编著. —北京:高等教育出版社,2004.8
ISBN 7-04-015226-6

Ⅰ.语… Ⅱ.张… Ⅲ.语言-测试 Ⅳ.H09

中国版本图书馆 CIP 数据核字(2004)第 067135 号

策划编辑 贾巍巍 **责任编辑** 艾 斌 **封面设计** 刘晓翔
版式设计 孙 伟 **责任校对** 艾 斌 **责任印制** 陈伟光

出版发行	高等教育出版社	**购书热线**	010-64054588
社 址	北京市西城区德外大街 4 号	**免费咨询**	800-810-0598
邮政编码	100011	**网 址**	http://www.hep.edu.cn
总 机	010-82028899		http://www.hep.com.cn
经 销	新华书店北京发行所		
印 刷	北京民族印刷厂		
开 本	787×1092 1/16	**版 次**	2004 年 8 月第 1 版
印 张	16.5	**印 次**	2004 年 8 月第 1 次印刷
字 数	300 000	**定 价**	26.50 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

内容提要

本书为告别纸笔考试，向计算机化的语言测试发展而作。全书从介绍当前语言测试计算机化的现状和发展入手，介绍计算机化的考试在考试现代化中的地位、性质和特点，计算机化考试的编制和实施，以及计算机的普及运用对考试的影响，并着重介绍目前国内外语言测试界常用于项目分析与题目等值的测试分析系统，使读者对当今语言测试界的现状与发展有较清楚的了解。书中提供获得这些测试分析系统的途径，对提到的每一种测试系统的功能做了详细的介绍，并给出最常用的适合中国教育国情的实例，便于读者结合本校和自己的实际对照、查阅、使用。

本书适合应用语言学专业和教育学专业的研究生、大学教师、学校教务部门以及考试、招生部门之人士阅读参考。

此为试读，需要完整PDF请访问：www.ertongbook.com

谨以此书献给我的博士导师，著名语言学家桂诗春教授！

前言

本书不再从理论上重复探讨语言测试中诸如信度与效度、常模参照性考试和校标参照性考试等基本概念，因为读者可以在桑戴克[1](E.L.Thorndike: 1903)的著作中找到这些最经典的教育测量学理论和定义，或者从桂诗春教授的专著[2]里获得对标准化考试、题库建设以及认知科学最全面的论述；本书也不准备追溯语言测试发展的历史过程，因为在伯纳德·思堡斯基[3](Bernard Spolsky：1975，1995)的专著中几乎汇集了全部有关当代语言测试发展的翔实史料，在李筱菊教授的专著[4]中也阐述了对当代语言测试发展三代体系的纵横观。因此，本书着重介绍和讨论当前语言测试中项目分析和等值技术的重要细节，以及最新和最常用的测试软件系统，力求脚踏实地，旨在与测试界同行交换看法。

众所周知，20世纪内语言测试经历了三次重大变革。在20世纪初，数理统计方法在考试中运用，其结果之一是教育测量学的诞生，随之而来的是在整个语言测试界兴起了客观性试题，这对当时的语言测试起到了极大的推动作用。到了20世纪50年代，电子计算机和光电阅读机在大规模考试中的广泛运用，促进了标准化考试的推广和普及。然而应当指出，这时的考试方法仍是传统的纸笔考试形式。到了20世纪末，随着测试理论的完善、计算机的普及、多媒体技术的运用和经济的增长，测试界已清楚地看到单一的纸笔考试弊大于利。那么纸笔考试所暴露出来的问题该如何解决?标准化考试还要不要继续推行?于是专家们开始在测试的方法和手段上下功夫，国外的TOEFL如此，国内的CET亦然。语言测试的方法和手段由此开始发生根本性变化，而这一变化竟来得如此之快，可以说是在向一个时代告别，即：向纸笔考试时代告别。同时意味着一个全新的时代的到来，即：计算机化考试时代的到来。作者认为，要全面实现计算机化的考试，项目分析和等值技术是当今测试界迈向计算机化考试的决定性一步，也就是说计算机化的语言测试应标志着工具的改进，从而使得考试方法得以根本改变。不仅如此，项目分析和等值技术对科学分析和评估整份试卷至关重要，对教学和命题等方面更具指导作用。这其中的每一步都是极为复杂而艰难的。

一般来讲，即使语言功底很扎实的人，也不见得会命题；命题专家们不

---

[1] E.L. Thorndike: *Educational Measurement*, 1903,1904,1975 等。

[2] 桂诗春：《标准化考试：理论、原则与方法》，广东高等教育出版社，1986年；《实验心理语言学纲要》，湖南教育出版社，1991年；“题库建设”载《题库建设：理论与实践》，国家教委考试中心主编，光明日报出版社，1991年，等等。

[3] Bernard Spolsky: *Measured Words*, 1995, Oxford University Press.

[4] 李筱菊：《语言测试科学与艺术》，湖南教育出版社，1997年。

见得了解项目分析；而对于从事语言测试、考试工作的人，如果其计算机、统计、测试理论功底不是很扎实，在遇到项目分析或题目等值时，往往也会采取回避态度。显而易见，这是因为项目分析和题目等值对测试理论、统计和计算机编程方面的要求较高，公式推导繁杂，术语概念甚多，不易推广。另一个原因是，国内讨论语言测试中项目分析的专著很少。除1985年桂诗春教授出版的《标准化考试：理论、原则与方法》一书中有专门章节详细介绍语言测试中的项目分析的方法之外，目前尚无同类专著。1990年6月，当时的国家教委考试中心曾召开国内第一次教育考试题库建设理论研讨会，会后由国家教委考试中心主编、光明日报出版社出版的题为《题库建设：理论与实践》一书中收集了20余篇论文，该书所收集的论文涉及题库中的试题命制原则、题库理论、测试方法、参数估算等，可以代表当时我国教育考试领域里的最高水平，同时也反映我国教育考试研究领域当时的状况。然而，涉及语言测试中的项目分析和等值技术的论文为数甚微。1996年9月，国家教委考试中心受国际教育评估协会执行委员会(IAEA)的委托，在北京召开了以“大规模考试的作用及相关问题”为主题的国际研讨会，来自世界33个国家和地区的160名代表参加并发言。会后同样由国家教委考试中心主编、外语教学与研究出版社出版的论文集[1]中收集了62篇论文，这些论文除涉及考试的社会问题、考试与社会、经济、教育和教学之间的关系、统计测量理论以及考试的具体技术方法等问题外，也开始出现关于计算机与多媒体技术对考试的影响等方面的研究论文。6年以后，全国大学英语四、六级考试委员会在上海召开以“语言测试与语言教学”为主题的国际会议，来自世界许多国家和地区的约200多名代表与会发言，会后由四、六级考试委员会编辑的论文集[2]中收集了31篇论文，作者介绍有关BILOG和PARSCALE的论文也收入在内。语言测试中的项目分析和等值技术方面的研究与应用受到测试界的关注。在语言测试向纸笔考试告别，向着计算机化考试发展的今天，语言测试中的项目分析和等值技术方面的研究与应用工作急待很好地深入开展。

本书专论语言测试中的项目分析和等值技术，但不可能解决语言测试研究发展中可能遇到的所有问题。本书力求突出实用性，使读者在较短的时间内，能够对当今语言测试研究中项目分析和等值技术的作用与应用，以及现状与发展有个清楚的了解，能够进入角色，了解一般语言测试研究所需的最基本的做法。因此，本书第一章简述经典测试理论与项目反应理论的异同，按国外最新测试理论的说法，列出项目反应理论和经典测试理论在十个方面的

[1] *The Effects and Related Problems of Large Scale Testing in Educational Assessment*，外语教学与研究出版社，1998，北京。

[2] *Language Testing and Language Teaching Conference Proceedings*，全国大学英语四、六级考试委员会，2002，上海。

不同，指出项目反应理论不是在经典测试理论基础上的继续和细化，而是另一门有着坚实基础的测试理论。同时对目前国内外常用的按项目反应理论编制的用于试题项目分析和等值的软件测试系统一一做了介绍，并提供获得这些软件的途径。第二章对国内正在使用的GITEST Ⅲ＋考试分析系统的研制、使用功能、用途、打印说明等做了专题介绍。第三章给出使用GITEST Ⅲ＋考试分析系统的实例和详解。第四章对国外正在使用的BILOG-W系统的使用功能、用途、输入文件、指令文件编排、指令句法结构、每一个指令的含义、用法，以及输出结果文件等做了专题介绍。第五章结合我国国情，给出最适合我国教育考试环境的使用BILOG-W的等值实例和详解。第六章对国外正在使用的PARSCALE-W考试分析系统的使用功能、用途、输入文件、指令文件编排、指令句法结构、每一个指令的含义、用法，以及输出结果文件等做了全面的介绍。第七章给出如何使用PARSCALE-W进行平行试题等值的实例和详解。第八章对GITEST、BILOG-W和PARSCALE-W的使用做出客观评价，让测试界同行了解使用这些软件系统的好处与不足。应当指出，这些实例和评价都是在多年的实际运用中总结出来的，部分数据为近年真实试测数据，凝结了编者多年来的大量心血和经验。

在最后的结束语中，作者展望了计算机化的语言测试今后的发展，认为我国计算机化的语言测试可以这么说："前途是光明的，道路是艰难的。"从托福考试等计算机化便可见一斑。这里有来自学术上的争论、技术上的限制、社会上的挑战、商业上的竞争，因而在今后一段时间内，计算机化的语言测试还不会完全占领测试舞台。但笔者认为计算机化的语言测试今后的发展在于用现代认知科学为指导。这样语言测试与现代认知科学、计算机科学的结合将表明一个学科的完全成熟。语言测试将不再是单纯的答对率和分数统计，而更是试题分析、情况反馈、多媒体技术的应用，是完全可以透过可观察到的分数去探索考生实际答题的全过程，即：人的内部认知心理的过程，给出的是一个综合考生答题全过程的一个能力值(θ)。从这个意义上讲，现代认知科学对语言测试的指导无疑会对今后计算机化的语言测试发展，对促进我国各类大规模考试计算机化起到极大的推动作用。

总之，世界已进入一个高科技时代，一切都要求科学性、系统性。这对语言测试界无疑提出了更高的要求。将来的考试要使全社会认可，使每一个考生信服，决非一张试卷、一张答题纸就能行的。因此，对于中国语言测试界，我们更应有强烈的紧迫感。如果我们现在还不能这样提出和认识这一问题，我们无疑将会落后于这个时代！

在本书即将付印之际，我要感谢我最敬爱的博士导师 —— 广州外语学院桂诗春教授，是他将BILOG引进中国，是他指引着我在语言测试这个领域里拼搏。同时我也要感谢GITEST的设计和编制者，李崴老师，在使用GITEST

的全过程中，他给予我及时的指导和帮助。美国SSI软件开发公司的Leo Stem先生和BILOG的设计和编制者之一，Bobert J. Mislevy教授，同时也是美国ETS特聘专家，对新版视窗版的BILOG使用给予了指导，美国ETS的Robin先生在PARSCALE的指令编程方面也给予了指导，在此一并表示感谢。

我还要感谢我的学校——解放军第一军医大学，她为本书的编写和出版提供了多方面的便利，给予了不懈的精神鼓励和物质支持。

高等教育出版社对于本书的出版给予了热情的支持和帮助，在此表示衷心的感谢。

作 者

2002年8月于广州

作者地址：广州第一军医大学外语教研室　510515

电子信箱：gzjohnzh@scut.edu.cn

目 录

第一章 语言测试的现状与发展 1

1 国外对经典测试理论和项目反应理论的研究 2

2 我国语言测试的现状 3

3 项目分析与试题等值的意义 4

4 项目分析的内容 4

5 试题等值的做法 4

6 试题等值所使用的模型 5

7 试题等值过程中对连接项题目的处理 5

7.1 连接项题目的数量要求

7.2 连接项题目的处理方法

7.3 连接项三种处理方法的利弊

8 用于项目分析和试题等值的软件 6

8.1 LOGIST

8.2 BICAL

8.3 BIGSTEPS

8.4 BILOG-W

8.5 BILOG-MG

8.6 MULTILOG

8.7 PARSCALE-W

8.8 RESGEN

8.9 GITEST

8.10 CET 考试成绩系统分析软件包

8.11 GGUM2000 系统

8.12 TESTFACT 系统

8.13 XCALIBRE 系统

8.14 RUMM 系统

8.15 RAGE 系统

8.16 CIPE 系统

9 计算机化的语言测试应具备的基本条件 10

9.1 题库建设

9.2 编程技术

10 理想的计算机化的语言测试应显示出的优越性 11

10.1 答题环境

此为试读,需要完整PDF请访问: www.ertongbook.com

10.2 考试时间
10.3 考试评分
10.4 试题的保密性得以加强
10.5 成绩报告准确无误
10.6 考试方法
10.7 操作方便，经济可行

第二章 GITEST III + 考试分析系统 15
1 GITEST III + 系统的功能 16
2 GITEST III + 系统的用途 16
3 GITEST III + 系统的数据输入文件 17
4 GITEST III + 系统的处理结果文件 18
5 GITEST III + 系统的处理结果文件说明 18
5.1 项目分析表
5.2 试卷总分析表和各大题分析表
5.3 试题难度与区分度双向汇总表
5.4 相关矩阵和因子分析矩阵

第三章 GITEST III + 考试分析系统应用实例 21
1 最常用的班级测试实例 22
1.1 题型
1.2 数据输入文件
1.3 数据输入文件的执行
2 GITEST III + 系统各选项输出文件的实例说明 23
2.1 按经典测试理论所产生的项目分析结果文件
2.2 按项目分析(单参数)分析所产生的以 -irt 为后缀的文件
2.3 按单参数模型等值所产生的以 -eqt 为后缀的文件
3 使用 GITEST III + 系统应注意的问题 44
4 小结 45

第四章 BILOG-W 软件系统 47
1 何为 BILOG？ 48
2 BILOG -W 软件系统的安装 48
3 BILOG -W 的主要功能 49
4 BILOG -W 的工作机制 50

5 BILOG -W 的使用方法 50
5.1 BILOG-W 的指令文件
5.2 BILOG-W 的数据输入文件
5.3 BILOG-W 的结果输出文件
6 BILOG-W 指令的句法结构 51
6.1 格式
6.2 指令和关键字
6.3 算术递增数的缩合形式
7 BILOG-W 指令的编排顺序 53
8 BILOG-W 指令的编写 53
8.1 Title 指令※
8.2 COMment 指令(即评述指令)※
8.3 GLObal 指令(即全局指令)※
8.4 BILOG 用于输入文件指定名称的关键字
8.5 SAVe 指令
8.6 LENgth 指令※
8.7 INPut 指令※
8.8 变量格式指令※
8.9 FORm 指令
8.10 TESt 指令※
8.11 CALib 指令
8.12 用于项目参数估算的 QUAdp 指令
8.13 PRIors 指令
8.14 SCOre 指令
8.15 用于能力参数估算的 QUAdp 指令
9 小结 76

第五章 BILOG-W 应用实例 77
1 最常用的班级测试实例 78
1.1 题型
1.2 数据输入文件
1.3 指令文件
1.4 指令文件的执行
1.5 输出结果文件
2 对同一份试题按各个部分进行分项测试的实例 88

2.1 题型
2.2 数据输入文件
2.3 指令文件
2.4 输出结果文件
3 使用 BILOG-W 进行平行试题等值的实例 102
4 使用 BILOG-W 进行题库扩充的方法 104
5 BILOG-W 系统常见错误信息 108
5.1 运行出错代码
5.2 指令编程错误
5.3 指令行及文件错误
6 小结 114

第六章 PARSCALE-W 软件系统 115
1 何为 PARSCALE? 116
2 PARSCALE-W 软件系统的安装 116
3 PARSCALE-W 的主要功能 117
4 PARSCALE-W 的工作机制 118
5 PARSCALE-W 的使用方法 118
5.1 PARSCALE-W 的指令文件
5.2 PARSCALE-W 的数据输入文件
5.3 PARSCALE-W 的结果输出文件
6 PARSCALE-W 指令的句法结构 120
6.1 格式
6.2 指令和关键字
6.3 算术递增数的缩合形式
7 PARSCALE-W 指令的编排顺序 122
8 PARSCALE-W 各指令的功能及用法 122
8.1 Title 指令※
8.2 COMment 指令(即评述指令)
8.3 FILes 指令(即文件指令)※
8.4 SAVe 指令
8.5 INPut 指令※
8.6 变量格式指令※
8.7 TESTi 或 SCAlEi 指令※
8.8 BLOCK 指令※
8.9 MGROUP 指令

8.10 CALib 指令※
8.11 QUAdp 指令
8.12 PRIors 指令
8.13 SCOre 指令※
8.14 QUAds 指令
8.15 COMbine 指令
9 小结 148

第七章 PARSCALE-W 应用实例 149
1 最常用的班级测试实例 150
1.1 题型
1.2 数据输入文件
1.3 指令文件
1.4 指令文件的执行
1.5 输出结果文件
2 使用 PARSCALE-W 进行平行试题等值的实例 153
2.1 题型
2.2 数据输入文件
2.3 指令文件
2.4 指令文件的执行
2.5 输出结果文件
3 PARSCALE-W 系统常见错误信息 231
4 小结 231

第八章 对 GITEST、BILOG-W 和 PARSCALE-W 的评价 233
1 数据处理类型方面 234
2 操作指令方面 234
3 功能选择方面 234
4 获得途径方面 235
5 GITEST III、BILOG-W 和 PARSCALE-W 功能对照表 235
6 GITEST III、BILOG-W 和 PARSCALE-W 所产生的项目难度参数比较 235
7 GITEST III、BILOG-W 和 PARSCALE-W 的指令使用情况 236
7.1 GITEST 输入文件

7.2 BILOG-W 指令文件
7.3 PARSCALE-W 指令文件
8 使用 GITEST III、BILOG-W 和 PARSCALE-W 的优点与不足 238

结束语 239
1 从认知科学的角度看目前计算机化考试 240
2 计算机化语言测试以及在我国的现状 240
3 计算机化语言测试所面临的问题 240
4 计算机化语言测试今后的发展 241

附录 243
附录一: GITEST III 项目难度参数表 243
附录二: BILOG-W 项目难度参数表 243
附录三: PARSCALE-W 项目难度参数表 246

CHAPTER

One

第一章

语言测试的现状与发展

此为试读，需要完整PDF请访问：www.ertongbook.com