



对外汉语/汉语国际教育研究生系列教材

总主编 崔希亮

语言测试概论：几个问题

张凯 著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS



教育部人文社会科学重点研究基地

对外汉语/汉语国际教育研究生系列教材

总主编 崔希亮

语言测试概论：几个问题

张凯 著



北京语言大学出版社
BEIJING LANGUAGE AND CULTURE
UNIVERSITY PRESS

© 2016 北京语言大学出版社, 社图号 16329

图书在版编目 (CIP) 数据

语言测试概论: 几个问题 / 张凯著. — 北京: 北京语言大学出版社, 2016.11

对外汉语 / 汉语国际教育研究生系列教材

ISBN 978-7-5619-4766-1

I . ①语… II . ①张… III . ①汉语—对外汉语教学—教学研究—研究生—教材 IV . ① H195.3

中国版本图书馆 CIP 数据核字 (2016) 第 276886 号

语言测试概论: 几个问题

YUYAN CESHIGAILUN: JIGE WENTI

排版制作: 北京创艺涵文化发展有限公司

责任印制: 姜正周

出版发行: 北京语言大学出版社

社 址: 北京市海淀区学院路 15 号, 100083

网 址: www.blcup.com

电子信箱: service@blcup.com

电 话: 编辑部 8610-82303647/3592/3395

国内发行 8610-82303650/3591/3648

海外发行 8610-82303365/3080/3668

北语书店 8610-82303653

网购咨询 8610-82303908

印 刷: 北京九州迅驰传媒文化有限公司

版 次: 2016 年 11 月第 1 版

印 次: 2016 年 11 月第 1 次印刷

开 本: 787 毫米 × 1092 毫米 1/16

印 张: 16.5

字 数: 333 千字

定 价: 49.00 元

PRINTED IN CHINA

“对外汉语 / 汉语国际教育研究生系列教材”

编委会

总 主 编：崔希亮

编委会成员：（以姓氏拼音为序）

崔希亮 孙德金 张 博 张 健

张 凯 张旺熹 张维佳

总 序

崔希亮

2010年初,由张博、孙德金教授提议,编写一套对外汉语/汉语国际教育研究生系列教材,推举我做总主编。我觉得这是一件值得去做的事就答应下来了。然后我们就组成编委会,开始确定选题,物色作者,并几次开会研究讨论,确定这套教材的编写思路,包括指导思想、教材定位、教材特色、编写体例等。北京语言大学出版社对这一选题也很感兴趣,从技术和资金方面给予支持。

编委会经过研究,确定了这套教材的八个板块,拟定了每个板块的内容,并初步确定了作者队伍。这八个板块是:

- 一、基础理论与方法板块
- 二、汉语语言学理论板块
- 三、语言教学理论板块
- 四、语言技能教学理论板块
- 五、语言要素教学理论板块
- 六、语言教育技术理论板块
- 七、汉外对比及跨文化交际理论板块
- 八、其他板块

根据这八个板块,我们拟分批推出30种教材:

- 1.《语言导论》
- 2.《语言学流派》
- 3.《语言学方法论》
- 4.《语言教学研究方法论》
- 5.《语言研究的实验设计》
- 6.《汉语概论》
- 7.《汉语简史》
- 8.《第二语言教学导论》
- 9.《第二语言教学设计概论》
- 10.《语言学的观念和方法》
- 11.《第二语言学习概论》
- 12.《语言测试导论》

13. 《第二语言教学史》
14. 《第二语言教学法研究》
15. 《专门用途汉语教学》
16. 《对外汉语教材导论》
17. 《汉语听说教学研究》
18. 《汉语阅读教学研究》
19. 《汉语写作教学研究》
20. 《语法及语法教学研究》
21. 《语音及语音教学研究》
22. 《词汇及词汇教学研究》
23. 《汉字及汉字教学研究》
24. 《语用、篇章教学研究》
25. 《语言教育技术研究》
26. 《语言教育资源研究》
27. 《汉外对比研究与教学》
28. 《跨文化交际研究与教学》
29. 《第二语言教学与文化》
30. 《第二语言教师发展与培养》

这个框架是按照学科的学理逻辑构建的，作者的选择也是根据学者们的学术背景和专业特长确定的。待各项工作展开之后，我们才发现这一工作并不简单。首先，教材的文本内容应该是深入浅出的，而要做到深入浅出实非易事；其次，这套教材应该考虑到趣味性、可读性，这又是一个更困难的任务，因为严肃的学术问题有的时候只能不苟言笑；第三，现在的教授们都很忙，忙着上课、带学生、讲学、做项目，甚至云游四海，很难静下来好好考虑教材的问题。由于有以上一些困难，有的作者表示难以完成任务，提出退出课题组，因此编委会经过讨论，决定从实际出发，适时调整板块内容和作者。

对外汉语教学/汉语国际教育是一个独立的学科，也是一个交叉的学科，同时它还是一个新兴的学科，这个学科涉及语言学理论、汉语语言学、语言教学和语言学习理论、语言教学技能、语言教学方法、语言测试、语言教育技术、汉外对比研究和跨文化研究等不同的分支学科。从学科分布上看，这个学科涉及语言学、教育学、认知科学、心理学、计算机技术等不同的领域，任何人都难于以一己之力贯通所有研究领域。因此整合学术力量、协力合作才是最佳选择。当前，开设语言学及应用语言学专业对

外汉语教学 / 汉语国际教育方向研究生课程的高等学校越来越多, 招收的研究生数量也十分可观, 尤其是专业研究生的数量增加速度很快, 而学生来源却是五花八门, 什么专业背景的都有, 师资力量准备不充分, 没有现成的合用教材。我们也知道这个学科既有的研究基础尚嫌薄弱, 但是编辑出版一套针对研究生的教材时机已经成熟。我们邀请了学界知名学者和汉语教学经验丰富的专家担纲, 推出这套“对外汉语 / 汉语国际教育研究生系列教材”, 可以填补这个领域的空白, 为对外汉语教学和汉语国际教育的研究生培养贡献绵薄之力。这也是学界期盼已久的事情。

这套教材的学科定位为语言学及应用语言学专业的对外汉语教学 / 汉语国际教育方向; 读者定位为硕士研究生指导教师和硕士研究生群体, 当然也涵盖汉语教师志愿者、孔子学院教师等; 内容定位为对外汉语教学 / 汉语国际教育的所有分支学科; 作者队伍定位为对外汉语教学与研究的一线教师和资深专家; 编写理念定位为理论与实践相结合, 学术性与可读性相统一。

在讨论的时候, 我们在以下几个方面达成共识, 即本套教材应该具备以下几个方面的特色:

(1) 以问题为导向设计章节, 目的是引导学习者尽快介入本学科研究的前沿课题, 深入思考各种各样的问题, 鼓励学习者自己动手寻找答案;

(2) 注意知识点和问题排列的科学性和系统性, 借鉴国内外学界最新研究成果, 求创新, 图新颖;

(3) 对问题的讨论力求深入浅出, 晓畅易懂, 不拘泥于成说, 诱导读者步步深入, 探及问题本质;

(4) 教材要有趣味性, 语言生动、例子经典, 学习者可以用来自学而不觉得枯燥乏味;

(5) 配合各章节提出的问题注重研究方法的讲解和阐释, 尽量给学习者留出进一步研究的空间。

科学研究是一项个体性非常强的行为, 所以本选题虽然是系列教材, 但是每一本教材的个性特征还是很鲜明的。尽管我们在体例上提出了要求, 但是不同的作者还是会把自己的印迹留在教材上。不管是内容的剪裁还是语言表达风格, 都与作者的研究领域和教学经验密不可分。

四年时间过去了, 作者们开始陆续交稿了, 我期待着学界的批评指正, 更期待着大家对这套教材提出完善意见和建议(反馈邮箱: gjhjxyj@blcup.com), 为日后的修订提供借鉴和参考。

2015年10月

第一章	绪论	1
第二章	语言能力问题	8
第一节	Bachman 语言能力模型批判	8
第二节	乔姆斯基的语言能力假说	31
第三节	乔姆斯基的 performance	47
第四节	语言测试应该测什么?	62
第三章	构想效度问题	64
第一节	构想效度概念的出现	64
第二节	批评	74
第三节	各种误解	76
第四节	构想效度是理论的有效性	82
第五节	1955 年以后效度理论的变化	100
第六节	现有技术的局限	117
第四章	项目反应理论 (IRT)	122
第一节	项目反应理论 (IRT) 简介	122
第二节	参数估计	133
第三节	项目信息函数和误差	142
第四节	不变性: IRT 的基石能撬动吗?	147
第五节	Rasch 模型的问题	179
第六节	IRT 和效度问题	188
第五章	联合测量	191
第一节	背景	191
第二节	基本内容及其错误	192

第六章 计量语言学和事件相关电位	207
第一节 计量语言学方法	208
第二节 事件相关电位	222
参考文献	237
后 记	251

第一章 绪论

语言测试属于应用语言学，而在应用语言学领域里，语言测试可以说是一门抱负不凡的学科。

在心理测量和教育测量的影响下，现代语言测试萌芽于 20 世纪初，“语言测试作为一个分立的活动的概念已初露端倪……在本世纪〔指 20 世纪——引者按〕的最初几年，可以看出，一场旨在测量人类差异的浩大运动已经开始”（Barnwell, 1996: 8-9）。1915 年，Frederick Kelly 发明了多项选择题；1919 年，Charles Handschin 编制出第一个标准化的语言测验（Barnwell, 1996: 9-15）。按照 Spolsky（1996: 354）的第二种划分，语言测试此时进入了它的“现代”时期。1961 年，Robert Lado 出版了他的 *Language Testing: The Construction and Use of Foreign Language Tests*，这是语言测试领域的第一部专著。

到了 20 世纪 90 年代，语言测试已经有了极大的发展，于是有人认为：

语言测试……从 Lado 时代应用语言学的配角，转变为今天〔指 20 世纪 90 年代——引者按〕在应用语言学中居主导地位的角色，其原因在于，它是用有明显活力的方法去面对语言学习和语言应用中的主要问题。在过去的 20 年中，语言测试一直使自己从只具有实践意义向既具有实践意义也具有理论意义转变。唯其如此，它便既获得了学院派的尊敬，又开始具备了解释力。（Davies, 1990: 71-72）

那么语言测试为什么能“获得学院派的尊敬并具备解释力”呢？按照 Davies（1990: 70-71）的说法，是因为语言测试“既要解决语言教学中的实践性问题，又要检验应用语言学所关心的核心问题：语言能力的性质、语言能力倾向、语言学习阶段的描绘”，因此，Davies（1990: 9-10）认为，语言测试有研究语言的作用。Bachman

(1990: 67-69) 则更具体地说, 语言测试可以用于研究语言能力的性质、语言处理、语言习得、语言丧失和语言教学等五个方面的问题。

从 20 世纪 90 年代开始, 语言测试的研究者就这样给自己的学科定位, 所以我们说语言测试可以算是一门抱负不凡的学科。

20 世纪 90 年代到现在, 又是 20 多年过去了, 我们怎样看待 Davies 等人给语言测试的定位呢?

Bachman (1990: 2) 曾说, “当前语言测验的研究和发展, 融合了几个领域的成果: 语言习得和语言教学研究、描述语言熟巧度和语言应用的理论框架以及测量理论”。既然语言测试研究融合了几个领域的成果, 那它或许也应该为这些领域做出一些贡献, 然而这样的东西似乎很少。Rogers (1995) 出版了一本 800 多页的书 *The Psychological Testing Enterprise: An Introduction*, 这是一部史论兼具的教材。他把书中所讨论的主要内容——其实就是心理测量以及相关领域的一些重大事件——列成一个表, 他称作 road map (路线图)。这个表长达十几页, 内容有几百项, 而语言测试方面的贡献只有一项, 即 Frederick Kelly 发明的多项选择题。

近一百年来, 我们语言测试领域一直在使用心理测量、教育测量、数学等领域创造的理论和方法, 而我们自己对其他领域几乎没有贡献。语言测试是应用语言学的一部分, Davies 等人甚至认为语言测试在应用语言学中居主导地位, 但我们给应用语言学乃至语言学提供了什么东西呢? 前文提到, 语言测试要解决语言能力的性质问题, 那么关于这个问题, 我们解决得如何呢?

20 世纪 60 年代, Lado (1961) 以及 Carroll (1968: 8) 提出了语言能力的“技能—成分”模型, Bachman (1990: 82) 对这个模型提出了批评:

为测量语言能力而建立的早期模型, 是把语言技能和语言成分结合在一起, 如 60 年代初 Lado 和 Carroll 提出的模型就是这样。这类模型区分了语言技能 (听、说、读、写) 和语言知识的组成成分 (语法、词汇、音位/正字法), 但没有指出各项技能和语言知识是如何联系的。各项技能是否只是知识成分在不同方式和通道的表现形式, 抑或它们之间有本质的区别, 这些都是不清楚的。

20 世纪 70 年代, 关于语言能力, Oller (1976、1979) 提出了他的“单一能力假说 (Unitary Competence Hypothesis, UCH)”, 他认为, 语言能力是一个单一的、不可再分的能力。尽管这个假说获得了相关分析和因素分析的支持, 但几乎从一开始, 就没什么人相信或愿意相信语言能力是单一的能力。有意思的是, 反对 UCH 的人则陷入一个悖论:

我们陷入一个这样的悖论：我们越是想测到不同的技能，并且测这些技能的工具越精良，这些技能之间的相关就越高，于是它们就越像是一个不可划分的单一技能。（Carroll, 1973: 11）

到了20世纪80年代，在Hymes（1972、1974）交际理论的影响下，Canale & Swain（1980）提出了一个语言交际能力模型。在这个模型的基础上，Bachman（1982）提出了自己的交际能力模型，该模型在80年代末又经过了修改和扩充（Bachman, 1990）。Bachman的交际能力模型声称通过了经验检验，语言能力的问题解决了！自Bachman以来，语言能力的研究在测试领域没有什么进展。然而现在我们发现，Bachman的做法有一些错误，这个模型根本没有通过检验（详见第二章）。

以上说的是语言能力理论方面的问题。自多项选择题、完形填空等题型出现以来，测试方式上至今也没什么大的发展。语言测试使用的题型仍然是以多项选择为代表的客观题和作文、口试等主观题。这两类题型各有各的问题。

我们先说客观题。客观题的基本形式包括两个部分：一部分是书面或口头的语言材料，这是给被试的刺激；另一部分是若干个书面的选项，被试从中选择一个他认为正确的作为他的反应。语言测试是以反应为中心的测量，我们是通过被试对语言刺激的反应去推测他的语言能力的。我们知道，接受语言刺激、做出语言反应，都是大脑活动的结果。然而，在使用多项选择题的情况下，我们无法知道被试大脑里的语言活动是什么样的，但我们却可以推断，在被试接受的刺激和他的反应之间，一些和语言能力无关的因素可能掺杂进来，这些因素可能会“污染”被试的反应，这些因素就是误差。大家知道，语言测试中的误差可能来自题目、评分员或评分系统，可能来自环境，也可能来自被试自己。我们只说来自被试方面的误差。假定我们使用的是四选一的多项选择题，假定有一个被试，他完全没有语言能力，但从概率上讲，他可以通过随机猜测答对四分之一的题目。我们又假定，有一个被试，他有一定的语言能力，他凭他的能力认真答题，也答对了四分之一的题目。用现在的技术，我们是无法区分这两个被试的。

至于主观测试，其缺陷就更是众所周知了，那就是评分的主观性。评分的主观性是客观存在的，现在，我们只能用增加评分员的办法去“抵消”这种主观性，但这并不是一个合理的办法（详见第六章）。

目前，语言测试在理论研究和技术开发方面都处于停滞不前的状态。我们也许认为语言测试是一门科学，至少我们希望语言测试这个学科能够成为一门科学。但也许我们忽略了，一个学科要想成为科学，那需要几个基本条件，归纳起来，这些条件就

是“四个基本”，即基本对象、基本现象、基本问题、基本方法。

科学不是什么东西都研究。在这个大千世界里，我们可以看到各种各样的事物。一个学科要想成为科学，它就要在这些事物中选择一种或一类作为基本的研究对象，而不是把很多不同的东西放在一起统统作为研究对象。例如，物理学中的力学只把运动着的大大小的物体当作研究的基本对象，这些物体大到天体，小到基本粒子。再如化学，它只以分子和原子层面上的物质结构为研究对象，大于分子或小于原子的物质它就不研究了。物理学和化学都不以社会、文化、历史等为研究对象。

波塞尔（2002：155-156）列出了科学研究中的“第一级规定”，其中第一条就是“本体规定”，“这类规定决定各个科学分支即学科的基本研究对象……另外还决定这些基本对象之间的一般关系”。波塞尔（2002：156）接着说：“没有一门学科可以离开这一规定……各个学科所特有的基本研究对象以及这些对象之间的基本关系，同时决定了该学科的结构。”

科学研究的目的是探索事物的规律，并使我们获得关于对象的真知识。不同的对象可能有不同的规律，例如，物体运动的规律就和遗传的规律不同。如果我们把性质不同的对象混在一起，那很可能就发现不了什么规律，反而会把事情弄得很乱。

在语言测试里，基本研究对象就是“语言能力”。对于这个判断，没有人会有异议，Lado、Carroll、Oller、Davies、Bachman 等人都明确地说语言能力是我们这门学科的研究对象。但是，如何理解以及怎样界定所谓的“语言能力”，不同的人可能有很不同的看法（详见第二章），在这个问题上，研究语言测试的人之间存在很大的分歧。

确定了基本研究对象，我们还要去寻找基本现象。一个或一类研究对象可能会呈现出多种现象，例如一个人，他在父母面前表现得像儿子，在妻子面前表现得像丈夫，在工作单位表现得像领导，而在医院则表现得像病人，这些都是这个人呈现出来的现象。和确定基本研究对象一样，我们也需要确定一个基本现象。一个特定的学科，把哪个现象当成其研究的基本现象呢？多伊奇（2002：144）说：

一个现象，如果只有深刻地理解了它，才能达到对世界的充分理解，则称这一现象为“基本”现象。当然，这个世界哪个侧面最值得去理解？哪个现象最深刻或最基本？对这种问题有各种各样的不同意见。有的人认为爱情是世界上最基本的现象，也有人认为如果某个人把经书背得滚瓜烂熟，那他就理解了一切值得理解的东西。我所讲的理解是用物理定律和逻辑、哲学原理表示的理解。深刻的理解是指那个更具有一般性、使表面上毫无关联的事实建立了更多联系、解释得更多而使用未解释的假设更少的理解。最基本的现象隐含在许多其他现象的解释

中，但是它们自己只能通过基本定律和基本原理得到解释。

就语言能力这个基本对象而言，我们可以看到它表现出的很多现象。例如，有人啰唆，有人寡言；有人声如洪钟，有人细声慢气；有人字斟句酌，有人信口开河。这些都是值得研究的基本现象吗？理解了这些现象算是对语言能力的深刻理解吗？或许我们应该在语言能力里寻找能用“物理定律和逻辑、哲学原理”去理解的现象？关于这个问题，我们将在第二章和第六章讨论。

确定了基本对象和基本现象，我们还要确定基本问题。什么是学科的基本问题？汪丁丁（1997：38-39）说：

基本问题就是贯穿于一个学科的全部历史并且推动着学科发展的那些问题。……基本问题在海德格尔看来根本就是不可能解决的，研究者只是不断地理解这些问题，不断地重新提出这些问题，从而不断地深化他们对整个理论的领悟。

汪丁丁这话也可以换一个说法：基本问题就是一个学科里最重要、最核心的问题，正是对这样的问题的不断研究，推动了学科的发展，而这样的问题，即便不是完全不能解决，也是很难解决的。在语言测试领域，语言能力（或语言能力模型）这个基本问题一直没有解决，至少没能很好地解决。在语言学领域，乔姆斯基在20世纪50年代提出了普遍语法理论，这个理论经过了若干个发展时期，直到20世纪末最简方案提出，普遍语法理论的问题也并没有彻底解决，但乔姆斯基及其学派的研究，推动了语言学的发展。

一个学科的基本问题为什么很难解决或不能解决，我们还可以从另一个通俗而简单的角度去理解：如果这个问题解决了，这个学科就没事干了，也就没有存在的理由了。

在我们这个学科里，我们试图解决的基本问题是：语言能力的本质是什么？这个问题，我们也将第二章和第六章讨论。

一个学科要想成为科学，还需要一个条件，那就是基本方法。皮尔逊（1892/1999：15）说，“形成科学的，不是事实本身，而是用来处理事实的方法”。在现代科学看来，这个基本方法就是数学方法。

冯·诺伊曼（1999：28）则说，“在现代经验科学中，能否接受数学方法或与数学相近的物理学方法，已经愈来愈成为该学科成功与否的主要标准”。

在语言测试（也包括心理测量和教育测量）领域，我们虽然一直在使用一些数学方法，做一些统计描述和统计推断，但我们做得还远远不够，我们还没有提出能用数学

方程来表示的关于语言能力的定律。关于这个问题，我们将在第三章和第六章讨论。

除了语言能力问题尚未解决以外，测试领域的其他一些问题也还存在着混乱和误解，例如效度问题，尤其是构想效度问题。测试中效度概念的提出，已经有 100 多年了。在这 100 多年的时间里，效度概念不断发展、扩充。1954 年，美国心理学会（APA）发布了《心理测量与诊断技术的建议执行标准》（简称“1954 年标准”），其中出现了一个新的效度概念，那就是“构想效度”。第二年，Cronbach & Meehl（1955）写了一篇文章，对构想效度概念进行了阐述。构想效度概念的出现，是测试领域的一件大事。Rogers（1995：472）认为，“构想效度的引进，可以看成是自比奈发明智力测验以来，测验领域中最为显著的变化”，它标志着测验“和历史传统与实用传统彻底决裂”，而且“至今仍以显著的方式改变着测验事业”。但有人说，效度研究方面“规范的理性进步却来得十分缓慢”（Wiley，1991：76），构想效度概念从出现时到现在，人们对它的误解，一直也没有消除。我们将在第三章讨论构想效度问题。

1952 年，Frederic M. Lord 写出了他的博士论文 *A Theory of Test Scores*，通常认为，这标志着被称为现代测量理论之一的项目反应理论（IRT）的正式诞生。经典测量理论（CTT）使用了大约 50 年后，人们发现它有一个很大的缺点，即测验和被试能力相互依赖。假如测验题目不变，能力水平高的被试会使这个测验的题目显得容易，而能力水平低的被试会使测验的题目显得难。这个事情的另一面是，假定被试不变，让他做一份难题，他的得分会低，而让他做一份容易的题，他的得分会高。这就是题目难度和被试能力之间的互相依赖。由于题目难度和被试能力互相依赖，我们就无法知道题目的绝对难度，也无法知道被试的绝对分数（能力）。Lord 的博士论文要解决的正是这个问题。2000 年，87 岁高龄的 Lord 去世，*Psychometrika* 杂志发布了讣告，讣告的作者是 Lord 的同事和朋友 H. Wainer 和 M. Stocking。讣告中说：

Lord 建立的统计模型，也就是项目反应理论，在实践上有两个重要成果。第一，它提供了一个估计题目难度值的客观方法；第二，它允许不同的被试使用不同的测验，而得到的分数仍然是可以比较的。

熟悉 IRT 的人都知道，这说的就是 IRT 的一个重要特性：参数的不变性。第一条说的是题目参数具有不变性，第二条说的是被试的能力值具有不变性。从 Lord 的博士论文出现，也就是 IRT 的出现到现在，60 多年过去了。在这 60 多年里，怀疑 IRT 不变性的人少，而承认不变性的人多。那么，IRT 的参数是变还是不变呢？我们将在第四章里讨论。

1964 年，Luce 和 Tukey 提出了一个新的测量理论——联合测量（conjoint

measurement), 关于联合测量理论的问题, 我们将在第五章讨论。

语言能力、构想效度、IRT、联合测量等理论问题都存在一些混乱, 这可能使我们的思路受到一定的限制, 同时测试技术的进步也受到了限制。本书最后一章将尝试着提出两个新思路。一个是用计量语言学的方法解决作文评分问题, 另一个是试图把事件相关电位技术引入语言测试。

第二章 语言能力问题

第一节 Bachman 语言能力模型批判

一、Bachman 的语言能力模型

Bachman 的语言能力模型可能是最著名、影响最大的模型。2012 年出版的 *The Encyclopedia of Applied Linguistics* 中有“Bachman”这个词条，其中说，在最近的 30 年里，Bachman 是语言测试领域最重要的理论家之一。Bachman 博士论文的内容是关于语言习得的，20 世纪 70 年代中期，他开始对语言测试产生兴趣。到了 20 世纪 80 年代初，他开始发表关于语言能力结构的文章。1982 年，他在 *TESOL Quarterly* 上发表了一篇文章（Bachman, 1982），内容是分析完形测验所测到的潜在特质的结构。Bachman 编了一个由 30 个完形填空题组成的测验，根据题目内容（删掉的词的性质），这 30 个题分为 3 类、13 组，其中句法 4 组，连接（词组或句子之间的连接）7 组，策略 2 组。这 13 组题目究竟代表了什么特质？可供选择的有三个模型：第一，13 组题目都属于一个单一的特质，这和 Oller 的 UCH 相当；第二，13 组题目分属“句法”“连接”和“策略”三个互不相关的特质；第三，“句法”“连接”和“策略”三个特质既能区分开来，又有一定的相关，这样，就构成了一个“一般因素加特殊特质模型”。经过拟合检验后，Bachman 接受了最后一个模型，这个模型是这样的：