

研究生教学用书

教育部研究生工作办公室推荐

语言测试科学与艺术

The Science and Art of Language Testing

李筱菊 著

湖南教育出版社

研究生教学用书

教育部研究生工作办公室推荐

语言测试科学与艺术

The Science and Art of Language Testing

李筱菊 著

湖南教育出版社

图书在版编目(C I P)数据

语言测试科学与艺术/李筱菊著.—2版.—长沙:
湖南教育出版社,2001.5

I.语… II.李… III.语言能力—测验
IV.H09

中国版本图书馆 CIP 数据核字(2001)第 05291 号

语言测试科学与艺术

李筱菊 著

责任编辑:陈 凯

湖南教育出版社出版发行

湖南省新华书店经销 湖南广播电视大学印刷厂印刷

787×960 16开 印张:33.5 字数:560,000

1997年3月第1版 2001年5月第2版第2次印刷

ISBN7—5355—2449—4/G·2444

定价:45.60元

本书若有印刷、装订错误,可向承印厂调换

序

作为一个外语老教师，出题目考学生是我常干的事，可是关于考试的那些道理我总是摸不清。我看过一些有关论著，还参加过专门讨论会，但那不过是浮光掠影，不能说没有收益，可是得到的知识不全面也不深。让我对语言测试有个完整的认识的，是李筱菊教授的新著——《语言测试科学与艺术》。

李老师这本书很好，可不容易用三言两语把它的特色描绘出来。这是学术专著，可又是一本教科书，一本科普读物。它深入讨论了语言测试工作，但为了初学者的方便，它首先讲了语言、语言教学和语言学习。论用途，命题的老师和应试的学生自然应该读它，可是负责搞测试的政府干部和学校领导人要是抽空把有关章节读一下，——不必念整本书，——便会觉得眼界大开，看清前进的方向。

我花了十多天读这本书，觉得很愉快，因为它出乎意外地浅。浅有两个原因。一是词语浅。书中引进了好些外国的专门术语，本来可能难懂，可是书中的译名既简又明。例如 proof reading 不译为“校对”，而译为“成段改错”，every nth word deletion 不译为“删掉每一个第 n 的词”，而译为“定距删词”。二是文体浅。书中文句结构严密，然而读起来好像是大白话，毫无钩章棘句之病，更无故弄玄虚之嫌。

在阅读过程中，我感到此书在理论方面有罕见的广度，又有惊人的深度。它回溯了几百年来各国各时代语言测试的历史，又综述了西方不同派别的主张和做法。它指出，就是同一个时期，同一个地方，甚至同一个试卷，那些考题也往往新旧杂糅，性质各异，有的合理，有的不合理或者无实效。它还指出，由于语言具有交际工具的特性，搞测试特别困难。你考语言知识吗？信度高，可是效度不高。你考语言运用的能力吗？效度高，可是信度不高。

书中讲理论讲得透，但是作者的视线主要集中在实际问题上。有许多事情，我们搞外语工作的人本来应该知道，可是多数不知，作者现在把真相都一起端出来了。例如美国的托福考试，不少人以为是很科学的

东西，可是本书告诉我们，托福那些考题是单考点、单句题型的，并不新，也不完善。国外出了一些 CALL 和 CALT 测试软件，使用了电脑，有些人以为很奥妙，很神奇，可是本书告诉我们，这些东西包装形式是新的，但命题方法却基本上是旧的。它还一针见血地说，我国的测试改革，尽管进行了十来年，有一个时期还很热心地搞，可是进一步退两步，必须痛下决心，进行到底。

一个人为什么花十来年功夫研究语言测试，写出一本如此深刻、精密、直率、大胆的书呢？是什么东西推动着作者锲而不舍，苦思苦想的呢？我想，这是因为她既有科学头脑，又抱严肃态度，又有菩萨心肠。一个人集学者、严师、慈母三种性格于一身，料想要她不写这本书也不成。是桃树就得开花，谁能下命令叫它憋住？

在欢喜、赞叹之余，我想提一点意见。Randolph Quirk 在编成了 *A Grammar of Contemporary English* (1972) 之后，随即写了简单明了的 *University Grammar*。为了便于一般读者，作者是否可以考虑写一本《语言测试简论》？

王宗炎

1995 年 12 月 30 日

于中山大学回春楼

前 言

书写完，觉得需要在前面说明三点。

第一点，说明这是本什么书，有什么用，适合什么读者对象。

书用中文而不用英文写，英语例子也都附汉译，是为了让英语以外其他语种的测试工作者、教师、学生都能用它，感兴趣的一般读者也能读它。

书本来就是以实际上课的教材为基础，整理、提高、系统化而成。适合做高等学校硕士、博士研究生或中、高等学校教师进修的教材。也可以做其他读者自学用书。书的每章附有作业及思考题。部分作业有参考答案。

其次，还可以用做语言测试工作者（尤其是命题人员）的工作手册。里面很多材料是多年培训命题人员用过的有效的材料。还有16年实际测试设计和命题工作总结出来的一条一条原则、要求和技巧。

同时，研究语言测试理论的人，也可以参考本书。它不属于纯学术理论著作，更不敢说有什么权威性的立论，但是书内对语言测试理论，尤其是对实践有指导意义的理论，作了一定广度和深度的挖掘，也算是笔者16年测试理论研究的总结。

第二点，大概说明一下书的内容。书共十一章。开头一章追溯语言测试的历史，最后一章展望语言测试的未来。中间九章都是讨论现代语言测试。九章里用一章（第二章）说明现代语言测试的理论。用三章（第三、四、五章）说明考试的建立及工作程序。然后用五章说明各种试题的命题。其中第六章先把试题及题型分类。第七、八、九、十章分别讨论四种试题的命题：单项试题（如语法或词汇填空试题）、综合项目试题（如完形填空试题）、接收性运用试题（读和听的试题）和产生性及综合运用试题（写和说以及听说读写综合试题）。这四章占了全书大半的篇幅。应该说，对命题的讨论比较透彻，立足于使学完本书的人能真正成为可用的命题人员。

第三点，要表示对所有过去和笔者一起搞过测试工作的同事们的由

衷的感谢。我之所以能在语言测试工作中取得所有这些经验，研究出所有这些理论，把它们写到这本书里，其实是很多人辛勤劳动的结果。首先感谢和我一起搞高考改革的桂诗春和李岚。然后感谢先后在广州外国语学院语言测试组和我共尝测试工作的甜酸苦辣的同事们：王初明、亓鲁霞、毛诚、源可乐、周宗诒、曾用强、邓旭东、王义、温宾利、季刚孟、许罗迈、王桂珍、张权、陈建平、肖南、祝丽仪，还有外籍同事 Martin Wedell, Michele Luxon, Sue Pitt, Nick Fletcher, Josephine Lewkowicz。当然还有在我的研究中给予帮助的国外众多的朋友们。对外籍朋友们，还要加一句：我的书用中文写，不能送给他们看，歉甚。

作者

1996年3月3日 于广州

目 录

第一章 语言测试的历史发展	(1)
1.1 纵观	(1)
1.2 横观	(3)
1.2.1 理论	(3)
1.2.2 实践	(11)
1.3 小结	(20)
第二章 现代语言测试的理论框架	(24)
2.1 语言测试科学和其他科学的关系	(24)
2.2 语言测试的实质	(30)
2.3 语言测试的根本要求	(33)
2.3.1 信度	(34)
2.3.2 效度	(37)
2.3.3 信度与效度的实现和矛盾	(43)
2.3.4 信度与效度矛盾的处理	(46)
2.4 小结	(47)
第三章 现代语言测试的实践(一): 建立考试	(50)
3.1 组织三方面的专业班子	(50)
3.2 关于考试的总体确定	(51)
3.2.1 确定考试类别	(51)
3.2.2 确定考试范围	(55)
3.2.3 确定考试的性质	(57)
3.3 设计试题结构	(70)
3.3.1 整体结构的设计	(70)
3.3.2 每道大题的定型	(73)
3.4 制定工作程序	(77)

3.5	制定考试大纲	(78)
3.6	小结	(79)
第四章	现代语言测试的实践(二): 考试工作环节	(82)
4.1	试题生产	(82)
4.2	考务准备和考试实施	(83)
4.3	主持考试	(84)
4.4	组织评分和评分	(85)
4.4.1	主观题评分	(86)
4.4.2	客观题评分	(92)
4.5	结果的统计分析	(94)
4.5.1	分数计算	(95)
4.5.2	分数检验和调整	(95)
4.5.3	对考试的分析评估	(96)
4.5.4	专题分析	(105)
4.6	反馈和总结	(106)
4.7	小结	(107)
第五章	现代语言测试的实践(三): 试题生产程序	(110)
5.1	试题生产的步骤	(110)
5.2	试题生产的两个组成部分	(112)
5.3	试题质量控制手段	(113)
5.3.1	质量控制手段之一: 审题员审题	(113)
5.3.2	质量控制手段之二: 实践检验	(113)
5.3.3	质量控制手段之三: 填写质量控制表	(114)
5.4	小结	(137)
第六章	现代语言测试的实践(四): 试题制作的总说明	(139)
6.1	试题分类	(139)
6.2	常用题型分类	(143)
6.3	试题指令	(148)
6.4	小结	(149)
第七章	现代语言测试的实践(五): A类试题的命题	(151)

7.1	非 MC 客观题型	(151)
7.2	MC 题型	(152)
7.2.1	格式	(152)
7.2.2	命题步骤	(153)
7.2.3	命题技巧	(158)
7.2.4	命题要求	(162)
7.2.5	题目分析与命题	(219)
7.3	小结	(226)
第八章	现代语言测试的实践(六): C 类试题的命题	(233)
8.1	完形填空试题	(234)
8.1.1	完形填空试题的格式	(235)
8.1.2	完形填空试题的效度	(247)
8.1.3	完形填空试题的命题要求	(260)
8.2	成段改错试题	(262)
8.2.1	成段改错试题的格式	(262)
8.2.2	成段改错试题的考点效度	(263)
8.2.3	成段改错试题的命题要求	(267)
8.3	小结	(270)
第九章	现代语言测试的实践(七): D 类接收性运用试题的命题	(274)
9.1	阅读试题	(274)
9.1.1	阅读试题的格式(题型)	(274)
9.1.2	阅读试题材料的选定	(277)
9.1.3	阅读试题的写题	(288)
9.2	听力试题	(337)
9.2.1	听力试题的格式(题型)	(338)
9.2.2	听力试题材料的选定	(339)
9.2.3	听力试题的写题	(343)
9.3	小结	(382)
第十章	现代语言测试的实践(八): D 类产生性运用试题及 E 类试题的命题	(395)

10.1	试题的性质	(395)
10.1.1	划定性质的两条界线	(395)
10.1.2	试题的效度、信度和可行性	(396)
10.2	试题设计(命题)	(401)
10.2.1	设计原则	(401)
10.2.2	具体设计	(408)
10.2.3	设计举例	(419)
10.3	小结	(429)

第十一章 语言测试今后的发展 (435)

11.1	科技发展对语言测试的影响	(435)
11.1.1	科技发展给语言教学和测试带来的新认识	(435)
11.1.2	科技发展在语言教学和测试中的新应用	(438)
11.1.3	新应用是否促进了新认识的实现	(441)
11.2	小结: 语言测试向何处去	(447)

附录 1 一九八三年全国高等学校招生统一考试题目 英语
(454)

附录 2 全国高等院校统一招生考试 英语科试题 (MET87 笔试)
(466)

附录 3 MET87 笔试项目分析表 (477)

附录 4 (University of London School Examinations Board) Certificate of Attainment in English, Graded Tests, Specimen Test, Language-Laboratory Version, Level Five (482)

附录 5 部分作业参考答案 (492)

附录 6 英汉术语及人名对照表 (506)

附录 7 主要参考书目 (517)

第一章

语言测试的历史发展

1.1 纵 观

语言测试，这里主要指外语测试，是随着外语教学而出现的。教学和测试有着密切的关系，要了解测试的历史，必须了解教学的历史。外语的教和学，起源于人类不同的语言社团之间开始有了交往的年代。学校里有正规的外语教学，并逐渐形成一套外语教学程式，却大约始于数百年前。从那时候到现在，外语教学不断发展和进化，衍生了名目繁多的外语教学方法与模式。外语测试的发展随在外语教学之后，也相应地演绎出种种方法与模式。

为了对外语教学和测试的发展史作一个概括的说明，需要从众多的外语教学和测试模式中，选择有代表性的加以分析。这里提出三个模式，与其说是某时某地的具体模式，不如说是外语教学与测试发展史中三大潮流的代表。这三大潮流，分别在先后的历史时期构成外语教学和测试的主流，对整个外语教学和测试有长远的影响。三个模式的特征，是外语教学和测试发展史里同范畴的方法与模式的特征的概括，因而具有广泛的典型性。它们还具有明显的区别性，代表着外语教学和测试的三条显然不相同的道路。

这三个模式，各自构成外语教学和测试的一个体系。体系，是理论和实践结合的整体。对典型体系的剖析，能够帮助我们获得对整体外语教学和测试发展过程的理解。这种理解自然不可能详尽，但重要的是能看清主干脉络。只有明了外语教学与测试历史发展的主干脉络，才有可能对今日外语测试的种种特征加以归类，对其走向作出判断。

外语教学与测试的三大体系，我们按其占主流地位的历史时期的先后，简称之为第一代、第二代、第三代体系⁽¹⁾。虽然称一、二、三代，代与代之间的起落交替，却并不界限分明。每一个体系，都由许多因素组

成。体系未成，有些因素早已存在；体系已衰，有些因素仍然延续。我们只能给每个大体系粗略地划出它成为明显的主流的一段时期，却并不排除实际上大大小小的体系与体系之间有并存，有交错；时期与时期之间有空当，有重叠。

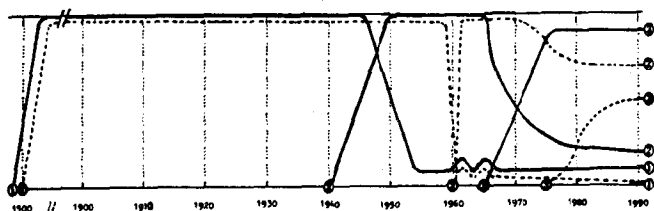
还有一种情况，虽然我们笼统地把三大体系称为外语教学与测试的三大体系，实际上它们在教学领域和在测试领域里的发展并不同步，就像轮唱里女声先唱起来，男声慢一拍才接着唱一样。外语教学与测试的发展，新理论一般都先在教学中开始贯彻。测试中的体现，由于种种主客观原因，往往要滞后十来年乃至数十年。同样道理，一种旧体系在教学领域里已经基本销声匿迹，在测试领域里却仍然可以长期保存其生命力。

另一种发展不平衡，当然就是洲与洲、国与国、区域与区域之间不可避免的不平衡了。因此，若要拿出一个外语教学和外语测试发展的纵观年代表，也就只能以外语教学与测试领先发展的地区（比如欧洲和北美）为依据了。在我国，与外语教学及测试相关的各种学科都起步较晚，整个外语教学及测试的进程，尤其是第二代及第三代体系的兴起及发展，比下图标出的年代至少要推后 30 年。

下面以三代体系为脉络，试绘一个外语教学与测试历史发展的粗略的纵观图。

【图 1.1】仅是三大体系的纵观，以第一代体系为起点。第一代体系，作为广义的外语教学与测试的传统模式，起始于何年何代，没有必要细加追究。我们需要知道的是，到了距今约四五百年前，在欧洲的学校里，人们教外语考外语，已经有了一个比较定型的、大家普遍遵循的方法，这就是我们称之为第一代体系的第一个外语教学与测试的主流。这个体系主宰外语教学与测试的时间特别地长，延续直至 20 世纪 40 年代。

【图 1.1】 语言教学与测试三大体系纵观略图



【图例】	第一代体系：	第二代体系：	第三代体系：
	教学①—①	教学②—②	教学③—③
	测试①…①	测试②…②	测试③…③

虽然一些属于第二代体系的观点在 19 世纪末便已有人提出,但是第二代体系之形成并占据主导地位却始于 20 世纪 40 年代。这个体系对整个外语教学和测试影响虽然巨大,它在教学领域中的统治却只持续了大约两个十年,至 60 年代中叶为止。值得注意的是,在测试领域里,直到 80 年代乃至 90 年代初,第二代体系仍然是世界范围内最大的主流。

20 世纪 50 年代,第二代体系就已开始受到挑战。60 年代后期,是第二代体系日渐失去垄断地位,而第三代体系开始崛起的年代。到了 70 年代后期,在西欧和北美,外语教学领域已完全是第三代体系的天下了。在外语测试领域,相对应的第三代体系,却并没有这样迅速的压倒优势。80 年代,在某些欧美国家,第三代的测试理论虽然已占了上风,在实践上仍和第二代模式平分秋色。就全球范围而言,第二代测试体系甚至可以说仍然占据着语言测试的广大市场。但是,第三代体系毕竟代表着历史发展的方向,80 年代后半叶以来,虽然从量上看仍未能成为主流,亦已形成一股能与主流抗衡,并逐渐迫使其让步的强大潮流了。

1.2 横 观

有了对三大体系发展的纵观,还得要有体系与体系相比较的断面观——横观。

为了对外语教学与测试三大体系进行剖析对比,我们从两个方面环绕五个关键点对每个体系加以分析。两个方面指理论与实践。理论方面要考虑的关键点是:语言观和语言学习观。实践方面要考虑的关键点是:内容的决定、教学与测试实施及评估。

1.2.1 理论

先分析三大体系理论上的异同,分析从第一个关键点语言观入手。

1.2.1.1 语言观 (theory of language)

语言观是一门很大的学问。我们不去探讨语言的抽象含义,我们只讨论直接作用于外语教学和测试的语言观,或称教学语言观及测试语言观。也就是说,语言教学,教的是什么?语言测试,测试的是什么?

语言观可以从内涵和外延两个方面去定义。

第一代体系,在教学上体现为传统语言教学道路 (the traditional teaching approach),在测试上体现为科学前语言测试 (pre-scientific testing)。在这个时期,语言学及其他有关学科尚未形成指导语言教学的系统

理论。主流外语教学基本上是一种凭经验、遵传统的教学。而外语测试，更说不上有任何科学理论依据。也就是说，第一代体系并没有自觉的语言观。然而从它的实践中，可以推论出它的语言观。按照传统外语教学体系去教语言的教师，虽然从不说语言是什么，却想当然地把它当做一门知识去教。这门知识主要包括语法知识、词汇知识和语音知识。懂不懂一门外语，就看你是否掌握这门外语这三方面的知识。语言测试，也就是测试这三方面的知识。这事实上就是把语言课和其他内容课、学科课、知识课等同起来。语言是一套知识，这就是第一代外语教学和测试体系的语言观的内涵。以此为内涵的语言观长期没有受到挑战，大家也就长期把语言当做知识教，当做知识测试。

到了20世纪40年代，人们开始感觉到，关于语言的规则的知识不等于语言本身，学了前者不等于学到后者。于是有人提出，教语言，教的应该是“语言”，不应该是“关于语言”（teach the language, not about the language）（Moulton 1961）。以美国语言学家布龙姆菲尔德（Bloomfield 1933）、弗里斯（Fries 1945）、拉多（Lado 1957）等为代表的结构主义语言学派（structural linguistics），第一次对语言的形式系统作了客观的、科学的分析。他们的论点是：语言是一套形式结构，一套符号系统。同时，受心理学的行为主义（psychological behaviorism）影响，他们认为学语言就是要获得操作这套符号系统的技能，用行为主义心理学的术语说，就是训练对刺激作出正确反应的一套语言习惯。结构主义语言学和行为主义心理学的结合，产生了结构主义教学道路（the structural approach）。同时，在测试方法上，也从心理学领域吸取了心理计量学（psychometrics）的科学方法，形成了心理计量结构主义测试（psychometric-structuralist testing）。这是外语教学和测试在历史上第一次以系统的科学理论为依据，创建了强大的外语教学与测试的第二代体系——结构主义体系。这个体系的语言观的内涵是：语言是一套形式系统。学语言考语言，学的考的正是操作这套形式系统的技能（skills）。教学语言观的内涵从知识转变为技能，体现了第二代体系对第一代体系的反动。

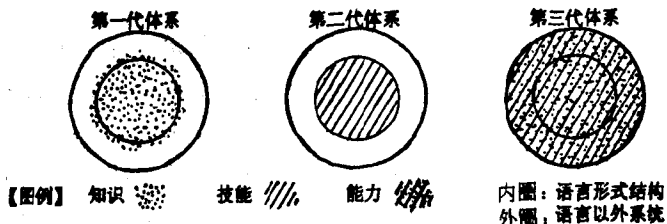
第二代体系最早受到的猛烈抨击来自生成语法学家乔姆斯基（Chomsky 1957, 1959）。他从心理学和语言学两方面驳斥了结构主义，提出了能力（competence）的概念。能力和技能是截然不同的概念。当一个人凭技能作出某种行为时，这种行为往往只是对外界刺激的习惯性的表面反应，是一种单纯由心理生理机制支配的行为。可是当一个人凭能力作出某种行为时，这种行为却是通过他的内在认知和情感力量而发出的

行为。争论点是：到底人的言语行为需要的仅仅是表面的技能，还是比技能要深得多的能力？正值此时，与语言学相关的种种学科——心理语言学、社会语言学、应用语言学、言语行为学说、语篇分析学说、第二语言习得学说等等——纷纷崛起并迅速发展，为寻求这个问题的答案提供了理论和实验依据。先有韩礼德 (Halliday 1970) 的语言功能学说，继而有威尔金斯 (Wilkins 1976) 的意念大纲理论。更直接的有萨维农 (Savignone 1972) 和魏多逊 (Widdowson 1972, 1973) 等，提出了新的语言教学体系——交际语言教学 (communicative language teaching)。这个体系认为，学语言不仅仅是学语音、语法、词汇知识，也不仅仅是训练操作形式符号的技能，而是获取人与人交际的一种能力。能力不排除知识和技能，但是也不等于知识加技能，而是把知识和技能包容进去的、性质起了变化的综合体。交际语言教学不停留于教知识，教技能，而要教能力。同时，在测试领域，相对应的外语测试第三代体系——交际语言测试 (communicative language testing, 亦称心理语言学—社会语言学测试——psycholinguistic-sociolinguistic testing)，提出语言测试所测试的，也应该是能力。

三代体系的教学及测试语言观，从内涵上分析，其基本区别正是：第一代体系认为教的考的应该是知识，第二代体系认为教的考的应该是技能，而第三代体系认为教的考的应该是能力。下面用简单的图来表示这个内涵上的区别。我们用三个圆圈分别代表第一代、第二代、第三代体系的语言观。每个圆圈的内容不同，表示每个体系的语言观内涵不同，不同的图示分别代表知识、技能和能力。

然后，我们再从外延去分析三代体系的语言观。教学和测试语言观的外延，指所教所考的“语言”范围有多大，包容些什么。既然第一代体系教和考知识，包括些什么知识？第二代体系教和考技能，包括些什么技能？第三代体系教和考能力，包括些什么能力？

【图 1.2】



传统语言教学和测试,教和考的是语音、语法、词汇知识。语音、语法、词汇是语言的三大形式系统。也就是说,这个体系的语言观的外延,主要限于语言的形式系统。在[图 1.2],内圈代表语言形式系统。第一个圆圈的内容,基本上都在内圈之内。不过也有少量内容超出了这个内圈,因为传统语言教学到了高级阶段,教师往往在传授语言形式的知识之外,兼讲一些背景知识、修辞知识、文学欣赏知识等等。传统的外语测试,到了高水平,除了考语言形式知识外,往往也附加上述这些内容。从图中可以看到,第一代体系教学和测试内容虽然略为超出内圈一点,所增加的也还是知识,内涵基本没有变,只是外延稍稍放宽了一点。

倒是到了第二代体系,语言观的外延反而收紧了。前面提到过,结构语言学派认为,学语言就应该学语言,不必学“关于语言”的知识。因此,他们的教学和测试,范围就严格限于语言形式系统。换言之,第二代体系语言观的外延和第一代体系基本上是相同的,都是语言形式系统(尽管第一代体系稍为超出了一点)。两代体系语言观的主要不同在于内涵不同。第一代体系把语言形式系统作为知识去教去考,第二代体系却把语言形式系统作为技能去教去考。

第三代体系,是对第一、第二代体系的反动,突破它们的局限性。突破不仅体现于语言观内涵的突破(知识、技能质变为能力),也体现于语言观外延的突破。乔姆斯基提出的能力的概念,固然对技能的概念说来是一个飞跃,但是他说出的能力也只是抽象的语言能力(linguistic competence)。很快就有语言学家指出,人要用语言进行交际,仅仅有抽象的语言能力是不够的。语言的运用,必须要放在具体交际情景中去。这就牵涉到很多超出语言能力的的能力。语言学家海姆斯(Hymes 1972)首先提出交际能力(communicative competence)的概念。这交际能力包括交际中所需要的语言能力及超出语言能力的的能力。第三代体系的教学及测试语言观以交际能力为语言的内涵,同时也就意味着语言观外延扩展了。再看[图 1.2],第三代体系不仅圆圈的内容不同,范围也从内圈扩大至外圈,把超出语言能力的的能力包括进去了。

第三代体系语言观的基本论点在形成过程中激起不少研究,环绕以下三个问题,引起过一些争论,也得出了一些结论,对语言教学和测试都有直接的关系。

一、交际能力大于语言能力,交际能力包含语言能力。这是整个第三代体系理论的大前提。正如上面所提的,20年前人们就认识到,一个人要会用一种语言,单凭对语言形式结构即语码本身的掌握,肯定是不