

外语测试学 导论

刘龙根 孙怀庆 著

吉林大学出版社

责任编辑 刘岩峰 封面设计 张沐沉

ISBN 7-5601-2032-6

H · 175 定价: 12.00 元

外语测试学导论

An Introduction to
Foreign Language Testing

刘龙根 孙怀庆 编著

吉林大学出版社

外语测试学导论
刘龙根 孙怀庆 编著

责任编辑、责任校对：刘岩峰

封面设计：张沐沉

吉林大学出版社出版
(长春市东中华路 37 号)

吉林大学出版社发行
长春市东方印刷厂印刷

开本：850 × 1168 毫米

1/32

1997 年 7 月第 1 版

印张：8.625

1997 年 7 月第 1 次印刷

字数：212 千字

印数：1 - 1000 册

ISBN 7 - 5601 - 2032 - 6/H · 175

定价：12.00 元

前 言

外语测试学这门溶现代语言学、应用语言学、教育测量学、心理学及计算技术等为一体的学科，其地位在国际上已确立多年。近年来，随着语言学理论研究的逐步深入、外语教学内容与方法的不断更新、教育测量评价手段的日益现代化以及相关学科的迅速发展，外语测试学的研究无论在广度还是在深度上都有了新的突破。然而，由于众所周知的历史原因，我国学者对外语测试的系统研究起步甚晚，直到 80 年代才逐步涉足这一领域。目前，我国的外语测试研究方兴未艾。

为了适应改革开放的需要，我国的外语教育事业有了长足的发展。然而，外语测试学研究的滞后无疑在很大程度上阻碍了外语教育的更快发展。其它各类外语测试不计，仅以全国大学英语四、六级统考为例，每年参考人数已达二百万人，其规模之大举世罕见。此外，许多有关外语教育甚或人事安排等方面的决策的制定都在一定程度上依赖于外语测试提供的数据。因此，确保外语测试所提供的信息具有科学性与可信性至关重要。为使外语测试科学可靠、公允客观，而又经济易行，加强外语测试学研究势在必行。

本书在广泛汲取国内外研究成果的基础上，旁征博引，系统探讨了外语测试学的理论、原理和基本概念；还本着理论紧密联系实际的原则，结合笔者外语教学与测试工作的实践，通过大量丰富的范例，具体阐述了测试各语言要素和各种语言能力的手段与方法，并且一分为二地分析了各种主要测试方法与手段的长处与不足、适用性与局限性，使之瑕瑜互见。本书对广大外语教师掌握语言测试的基本理论和命题技巧能够发挥应有的作用，并且对各类英语学习者在准备参加相应水平的英语

测试时了解测试的性质、类型等有一定的裨益。

外语测试学的理论与实践都在继续不断地发展。笔者所见难免有疏漏之处，本书旨在抛砖引玉，把我国外语测试学理论的研究推向新的高度，诚望广大外语教学界的专家和同仁指正。

最后，本课题作为吉林大学笹川良一优秀青年教育基金项目，承蒙有关各方的关怀与支持，得以如期顺利完成；尤其是吉林大学出版社与社科处为本书的出版做出了辛勤的努力，在此谨致谢忱。

编著者

1996 年 12 月

目 录

第 1 章 绪论	(1)
1.1 应用语言学与外语测试学	(1)
1.2 外语教学与外语测试	(2)
第 2 章 外语测试理论发展的四个历史时期	(3)
2.1 前科学时期	(3)
2.2 心理测量—结构主义时期	(5)
2.3 心理语言学—社会语言学时期	(11)
2.4 交际语用时期	(20)
第 3 章 外语测试的类型	(27)
3.1 速度测试与难度测试	(27)
3.2 主观性测试与客观性测试	(29)
3.3 常模参照性测试与标准参照性测试	(30)
3.4 五种基本测试形式	(32)
第 4 章 外语测试的评价标准	(41)
4.1 效度	(41)
4.2 信度	(45)
4.3 可行性	(49)
4.4 可接受性与后效潜势	(50)
第 5 章 语法测试	(52)
5.1 测试内容	(52)
5.2 测试题型	(53)
第 6 章 词汇测试	(69)
6.1 词汇的选择	(69)
6.2 常用题型	(72)
第 7 章 听力测试	(95)

7.1 概述	(95)
7.2 测试内容	(97)
7.3 测试类型与形式	(100)
第8章 阅读测试	(132)
8.1 阅读理解与阅读测试	(132)
8.2 测试文章的选定	(134)
8.3 测试题型	(138)
第9章 写作测试	(175)
9.1 概述	(175)
9.2 命题	(177)
9.3 评分	(201)
第10章 口语测试	(217)
10.1 概述	(217)
10.2 测试内容	(219)
10.3 测试形式	(220)
10.4 评分	(234)
第11章 外语题库建设与项目反应理论	(245)
11.1 外语题库建设	(245)
11.2 项目反应理论	(246)
第12章 测试数据的统计分析	(251)
12.1 集中趋势统计	(251)
12.2 离散度统计	(253)
12.3 题项分析统计	(255)
主要参考文献	(261)

第1章 绪 论

1.1 应用语言学与外语测试学

随着语言学理论研究的逐渐深入,外语教学内容与方法的不断更新,教育测量评价手段的日益现代化以及相关学科的迅速发展,外语测试学的研究无论在广度上还是在深度上都有了新的突破。它已逐步在应用语言学研究中立起核心地位,并且日益成为其重要的学科分支。较之应用语言学其它的实践活动,语言测试实际上历来是语言学原理与方法得以应用的主要活动。从测试内容的抽样、测试形式的择定、直至测试结果的评价与报导,无一不涉及对语言与语言学习本质的认识,因而都具有应用语言学理论的意义。不言而喻,现代外语测试必须以先进的语言学理论为基础。同时,语言学研究中遇到的许多问题往往需要在语言测试中寻觅答案。语言学关于语言的本质、语言学习的性质以及语言能力之实质的各种理论假说一方面经常须在实验中借助于测试手段加以检验,从而得到扬弃、修正和发展,另一方面又不断地影响着对语言测试性质的认识和对不同测试方法的选择与运用。在这方面,颇为典型的例子是关于语言能力两种观点的对峙导致语言测试形式上的分歧(详见第2章)。这两种观点便是所谓的“可分能力假说”(divisibility hypothesis)和“整体能力假说”(unitary competence hypothesis)。按照前一种假说,人们的语言能力是可分的,旨在考核语言使用不同方面的测试所取得的结果将能区分这些不同方面。这种“可分能力假说”为结构主义语言测试学家与一些交际语用测试

学家所推崇。相反,根据“整体能力假说”,语言能力是一个不可分割的整体,用于考核不同语言结构的测试不会产生互不相同的结果;这些测试结果之间所存在的差异类似于旨在考核同一语言结构所进行的那些测试所获结果之间的差异(Porter, 1983)。正是基于这种观点,心理语言学—社会语言学测试学家提出了综合测试法。

1.2 外语教学与外语测试

尽管有人将测试称作“a necessary evil”(难免的坏事),外语测试在外语教学中仍是个不可缺少的重要环节,而且经常还是一种行之有效的教学手段。通过适当形式的外语测试,教师可以掌握学生的学习潜能,了解学生学习进步程度以及所存在的问题,评价学生的外语水平和实际应用外语的能力,检查教学大纲与教学计划的落实和完成情况,考查教与学的实际效果。科学合理的外语测试还能激发学生学习外语的兴趣,增强学习的动力。当然,如果外语测试质量低劣或使用不当,同样也会造成极大的负面影响。如何有效地发挥外语测试的积极作用,最大限度地克服其负面效应,正是外语测试学需要认真研究的重要课题。

第2章 外语测试理论发展的四个历史时期

语言测试方法不仅反映语言教学的目标,而且体现了关于语言和语言学习本质的理论假设。本世纪以来,随着语言学习目标的改变和各种有关语言与语言学习本质理论的提出,语言测试的理论与方法也经历了四个既有区别、又相互重迭的历史发展阶段,即所谓的前科学时期、心理测量—结构主义时期、心理语言学—社会语言学时期和交际语用时期。

2.1 前科学时期(the prescientific stage)

前科学时期指的是本世纪50年代之前较长的一个时期,其影响一直延续至今,并在一些语言教师的教学与测试中得到体现。所谓前科学,尽管似乎从未有人给以明确的定义,也许包括两种不同的含义。一方面,它表明当时的外语教学与测试方法缺乏与某种具体语言学理论的联系(cf. Whiteson, 1981: 346)。另一方面,它或许显示这个时期的外语测试缺少更为科学的测试方法所特有的心理测量标准。从这两种意义上说,这个时期的外语测试也许可以被冠以“前科学”字样。但是,毋庸置疑,这一时期的测试手段是与当时外语教学的目标相一致的。

在“前科学时期”,即二次世界大战以前的很长一个阶段里,外语教学的主要目的是为学习者提供打开外语文学宝库的钥匙,帮助他们通过掌握外语能够阅读、欣赏外国文学原著。围绕这一教学目标,外语课堂主要采用语法翻译法和以阅读为中心的教学方法。毫不奇怪,这些教学方法把教学的重点几乎全部

置于书面语的学习掌握之上,即口头交际能力的培养受到严重的忽视。它们旨在通过对母语与目的语互译的严格训练,使学生能领悟目的语的语法规则,并且培养他们准确使用目的语进行写作的能力。同时,语法翻译法还旨在通过翻译实践,培养学生理解外语篇章,进而赏析所读作品中的文

及文化内涵之能力。与这些外语教学方法相适应,语言测试当时采用的形式通常与课堂练习形式相差无几。最主要的方式是要求学生将一篇相当长的文章译成外语或母语(Farhady, 1979)。此外,还有的测试要求学生做书面问答题、完成句子或命题作文等。所有这些测试形式都不同程度地服务于某种文化或语法教学目的。

前科学时期外语测试中的试卷主要由课堂任课教师自行编制,并没有语言测试专家参与其中。一种心照不宣的假设是,凡是胜任外语教学的教师也同样胜任于外语测试工作(Clark, 1983)。结果,前科学时期采用的测试方法从实质上看缺乏对测试效度与信度的关注。

针对前科学时期的外语测试未充分重视测试的信度与效度这一现象,后来的批评家曾进行过猛烈的抨击。譬如,著名的语言测试学家 Lado (1961)就认为将翻译和写作用以考核学生对外语的掌握程度,其效度和信度令人怀疑。他觉得翻译能力并不与外语的听、说、读、写能力成正比,因为翻译能力是一种特殊能力,这种能力并不能与前四种能力直接互为因果关系。同翻译一样,短文写作作为一种测试形式也受到了质疑。理由是,具备了写出一篇好作文的能力并不一定能确保学习者具备使用目的语进行口头或书面交际的能力。此外,翻译与写作测试往往都会遇到信度问题。同一篇译文或作文由不同的教师来评,评分结果也许会大相径庭,这种情况绝非鲜见。然而,尽管翻译与写作作为测试手段遭到了诸如上述的种种抨击,我们也决没有理由走向极端,彻底剥夺它们在外语测试中的一席之地。

目前,国内的一些外语教学大纲都将培养学生具有一定的翻译能力作为教学目标之一。这种教学要求理所应当要有相应的测试手段加以检验。为此目的,正如 Lado (1961:32)本人所指出的那样,一项经过合理设计、评分方法更加客观的翻译测试完全可能起到无以替代的作用。这一论点在近年来的一些专题研究中也引起了共鸣。一些学者(如 Titford, 1983; Stansfield et al, 1992)注意到了翻译作为一种教学手段应当重新为人们所重视,并且提出应当重新认识翻译在外语测试中的作用。

2.2 心理测量—结构主义时期 (the psychometric-structuralist stage)

心理测量—结构主义语言测试模式尽管形成于本世纪五、六十年代,其影响却经久不衰,一直延续至今。基于心理测量—结构主义理论模式所产生的测试方法(尤其是分离式多项选择题型)在各种外语测试中仍然广为使用。众所周知,语言测试方法不仅从一个侧面反映了语言教学的既定目标,而且也在一定程度上体现了人们对语言与语言学习本质的理性认识。故此,为了使心理测量—结构主义模式(以下简称心理测量学模式)更有效地服务于我们的外语教学与测试,避免盲目地照搬滥用,有必要对其理论基础和实践方法加以系统的探究,从而进一步丰富我国的外语测试理论。同时,通过客观地剖析其适用性与局限性,我们才能使之瑕瑜互见,进而在同其他测试模式与方法的结合使用中取长补短、相得益彰。

心理测量学时期的外语测试区别于以往测试方法的主要之处在于它一方面首次以当时盛行的关于语言与语言学习的理论为基础,另一方面有测试学专家的直接参与。究其理论渊源,行为主义心理学和结构主义语言学皆留下了不可磨灭的烙印。在行为主义心理学的影响之下,语言学习被视为一种刺激—反应的过程和通过积极强化进行习惯培养的过程。同时,按照结构

主义语言学的观点,语言教学的重心应置于语言的形式结构之上,而语言形式可以独立于所含内容加以系统描述和处理。因此,复杂的语言体系在不同的结构层面上被切分成孤立的语言成分。

心理测量学测试模式的兴起在很大程度上获益于著名的美国语言测试学家拉多(Lado, 1951-60)的奠基性研究。拉氏外语测试方法的理论基础即为结构主义语言观与行为主义语言教学观。他的语言测试理论假定,语言是一种交际习惯体系(Lado, 1961:22)。因而就外语学习而言,母语习得中养成的某种习惯必然以一定的方式“迁移”。“迁移”有正负之分:本族语与目的语之间存在相似之处时产生正迁移;两种语言存在差异时,由于学习者生搬硬套母语知识而导致负迁移。以初学英语作为外语的中国学生为例。他们在学习英语陈述句时几乎毫无困难,因为这种英语句式同汉语句式的构成类似,通常均为主语+谓语(+宾语)的结构。然而,他们在学习英语疑问句时却颇伤脑筋,因为这种英文句式常常要求将相应陈述句中的主语与功能词颠倒,而汉语则无此要求。两种语言中类似这种差异的例子不胜枚举。故此,依据结构主义语言学的对比分析原则,通过系统比较外语与母语之间相应结构的异同,我们能够预见并描述外语学习中的难点。这些难点一经如此确定和描述,就应当成为外语教学的重点,从而帮助学生最大限度地降低负迁移。同理,旨在评价学生外语掌握程度的语言测试也只有针对这些难点进行才能奏效。在这一点上,拉多本人的话真可谓一语中的:“我们力图考核这些学习难点。其理由在于,掌握了这些难点亦即掌握了该语言。我们明确地说,考核难点就是考核语言。”那么,如何才能最为有效地考核这些外语学习中的难点呢?

心理测量学语言测试模式的倡导者们坚持认为,诸如翻译和作文等传统的外语测试方法无以直接用于考核学生对由对比分析所揭示的语言难点的掌握程度。为此,需要研究设计出题

目更加具体、针对性更强的题型。正是为了满足这种需要,拉多的巨著《语言测试》于1961年问世了。在这部专著中,他借助于丰富翔实的范例,具体阐释了测试语音、语调、词汇、语法等语言成分和听、说、读、写等语言技能的各种方法。无论测试何种语言成分,命题者都对试题的语言结构严格加以限定。除了构成难点的语言,其他部分或完全排除、或控制在最低限度,以便为测试的要点提供必要的语言框架(Clark, 1983:432)。这种由拉多等人创立的“一道试题测试一个语言点”的方法现在通常称作“分离式”测试法。分离式测试中最常用的题型当数多项选择题。这种题型在各种常见的大规模标准化外语考试中被普遍采用。以著名的“托福”考试为例,自1963年启用以来,尽管近年局部有所变更(如增加了写作内容等),但是,多项选择题始终占据主导地位。下面仅列举几道典型例题,藉此,可对这种考试方法的性质与特点略见一斑:

1. I wish you _____ me of their arrival earlier.
*A. had informed B. have informed
C. informed D. inform
2. I simply don't know what you are driving _____.
A. to B. on C. by * D. at
3. They intend to commence the research project as soon as possible. *
A. stop B. cancel * C. start D. do
4. At no time you should entertain the hope that you will be
A * B C
able to wash your hands of the business.
D

从这几道例题中不难窥得“一道试题测试一个语言点”的分离式测试原则之真谛。例(1)针对的是 wish 之后宾语从句中谓语动词的虚拟结构,例(2)和(3)考核词汇,前者测试考生对 drive at 这一短语动词的掌握情况,后者通过选同义词测定考生对 com-

mence 一词的认知程度,而例(4)通过要求考生鉴别错误项考核学生有关英语倒装结构的知识。

目前,由于“托福”、EPT 等大规模标准化考试的影响,我们很多的外语课堂考试乃至教学练习也大量采用分离式多项选择题,其负面效应已日益明显地暴露出来。因此,在了解了其理论基础之后,还必须对这种测试模式的长处与不足和实际使用中的利弊得失有一个清醒的认识。

心理测量—结构主义模式的分离式测试方法区别于传统测试方法的又一特征在于,它首次明确地将效度与信度作为评判外语考试的基本标准。语言测试中有史以来第一次提出了有关考试是否真正考核了想要考核的内容以及这种考试的结果是否科学可信的问题。有关的测试专家汲取了当时居主导地位的言语学与语言学习理论,研究设计出了客观性测试题型和分析型评分方法。他们的工作原则植根于这样一种信念,即学生的语言水平能够通过客观性考试有效而可信地加以考核。这样,分离式测试法就应运而生了。分离式测试中最为典型(尽管不是唯一)的形式是多项选择题测试。

分离式试题自开始在外语测试中投入使用以来,就不断地受到各方面的褒贬毁誉。在其鼎盛时期,这种测试方法不仅在语言测试中独执牛耳,其影响几乎遍及整个教育领域,并且成为一些非英语国家考核英语作为外语或第二语言学习者英语水平的唯一科学可靠的测试方法(cf Farhady, 1979)。如前所述,分离式测试法目前在世界一些地区仍然有着相当广泛的市场。毋庸置疑,在考核学生对目的语语言体系具体要点的掌握程度上,分离式测试题较之其它题型具有许多优越之处。譬如,分离式试题使词法测试(包括词缀与词干的测试)简便易行。同时,它在一定程度上也为评价学生正确应用时态、语态、内嵌规则等的语法能力提供了有效的途径。正如前节例题所示,分离式试题一般都具有其他题型难以比拟的若干明显的优点。除了考试易

于实施,阅卷评分可借助于机器而省时省力外,这种考试评分客观公允,试题内容具体、针对性强,且覆盖面较宽。分离式试卷经过适当加工还可用于帮助学生自学,或者对其语言知识学习的进步情况进行自测。分离式测试方法尽管具有上述诸多优点,近年来却不断遭到外语测试学界的抨击与质疑。其中最严重的挑战是针对其理论基础发起的。正如前文所指出的那样,心理测量学语言测试模式基于的理论假设是语言体系可以在不同的结构层面上被切分成一个个语言成分,而对这些成分的掌握也就等同于对该语言的掌握(见 Morrow, 1979)。正如许多学者(如 Porter, 1983; Davies, 1990)所指出的那样,这种理论假设并不能完全成立。这些学者认为,现代语言学研究充分表明,语言具有生成性和冗余性等基本特征。因此,语言成分的简单相加并不完全等同于语言体系的总和。类似这种质疑,似乎在某种程度上削弱了分离式语言测试法的效度,至少是就其构卷效度(construct validity)而言。

除了对其理论假设进行抨击之外,一些学者还针对分离式试题(如多项选择题)的测试形式和测试内容提出了质疑。首先,许多学者(见 Walker, 1983; Weir, 1990; 等等)指出,分离式测试忽略了考核实际语言运用能力。这是因为,这种测试要求应试者所完成任务纯属人为所致,缺乏交际真实性。除了在应试场合,考生在其他语言使用场景中是不可能遇到类似任务的。其次,试卷上无论使用什么样的语言,实际上皆为命题者的语言,而并非应试者的语言,因为应试者只需在若干个选择项中鉴别出正确答案,而无须用自己的语言将答案内容表述出来。然而,能够辨别正确答案并不总是意味着有能力用自己的语言正确地表达这一答案。对某个概念似懂非懂的考生可以凭借排除法,剔除他所能甄别的错误选择(即干扰项)而确定正确答案。事实上,这正是一些教师在应试训练中面授机宜般地传授给学生的解题窍门。由此而言,这种考试的特性也许在一定程度上