

中学教师继续教育丛书

英语测试的原则与方法

杨钟琳 编著



杭州大学出版社

中国翻译理论教育丛书

英语测试的原则与方法

陈建明 编著



杭州大学出版社

英语测试的原则与方法

杨钟琳 编著

杭州大学出版社

(浙)新登字第 12 号

英语测试的原则与方法

杨钟琳 编著

*

杭州大学出版社出版

(杭州天目山路 34 号)

*

浙江省新华书店发行 杭州兴业印刷厂印刷

787×1092 毫米 1/32 8.5 印张 178 千字

1992 年 8 月第 1 版 1992 年 8 月第 1 次印刷

印数: 0001—5000

书号: ISBN 7-81035-239-3/H·011

定 价: 3.70 元

出版说明

为了配合中学教师继续教育的需要，在上海教育学院院长张家祥教授倡导下，北京、天津、上海、浙江、四川、辽宁、广东、陕西、湖北、江苏、武汉、杭州等十二省市教育学院联合发起，决定编写出版《中学教师继续教育丛书》。国家教委师范教育司对这项工作给予了热情的支持和具体的指导，并委托我社承担丛书的出版任务。

中学教师的继续教育，是指中学教师按现行学历规定合格后的非学历的着重于提高政治、业务、教育教学能力的教育，因此《丛书》不能照搬照抄大学本科生段和研究生段的课程内容，也不能等同于现行中学课本的备课资料。

《丛书》应具有科学性、先进性、适用性、针对性，应致力于学科前沿知识与基础知识同中学教育实际的结合，教育科学与心理科学同中学学科教育实际的结合。

《丛书》书目大体分为三类：一、思想政治教育和道德修养类；二、专业知识的拓宽和更新类；三、学科教育和学科心理类。其中以第三类为重点。

《中学教师继续教育丛书》设编委会，负责领导和组织《丛书》的编写工作。上海教育学院院长张家祥教授任主编，杭

州大学校务委员会副主任金锵教授任副主编。编委单位有下列院校：

天津教育学院

四川教育学院

湖北教育学院

杭州教育学院

广东教育学院

武汉教育学院

北京教育学院

陕西教育学院

辽宁教育学院

浙江教育学院

江苏教育学院

杭州大学

《中学教师继续教育丛书》的编写和出版，是一项新的探索性的工作，需要从事中学教师继续教育的同行和中学教育界的广泛支持。我们热切地期待着大家的批评、指正，以便把这套《丛书》编得更好，为提高中学教育质量，发展中学教育事业，贡献我们的一份力量。

杭州大学出版社

1991年8月

总 序

由京、津、沪、浙等全国十二省市教育学院协作编写的《中学教师继续教育丛书》陆续出版了。在更新教育观念、深化教育改革的今天，这套丛书的问世是很有意义的。

国家振兴，教育为本；教育振兴，教师为本。能否建设一支思想品德素质和文化业务素质精良的师资队伍，关系到社会主义教育事业的成败。而要加强师资队伍的建设，就得采取必要的措施，使他们能结合工作的需要，不断地再学习、再进修、再提高。

随着大部分中学教师逐步达到现阶段国家规定的合格学历，教师培训工作的重点必将有步骤地转移到开展继续教育上来。这种继续教育是指对已达国家规定学历的教师进行以提高政治思想素质和教育教学能力为主要目标的培训。它包括职务培训、新教师见习期培训、骨干教师培训和对部分骨干教师提高学历层次的培训等方面。做好这项工作，对于建设一支能够坚持社会主义方向，品德高尚，素质优良，结构合理，适应我国教育事业发展需要的教师队伍，有着十分重要的作用。

开展继续教育，不能没有教材。但我们的国家地域辽

阔，人口众多，各地师资队伍建设的客观条件和实际需求很不一样，这就需要从实际出发设置相应的课程，编写不同的教材。这次，一些起步较早、条件相仿的教育学院，根据已有的实践，发挥群体的优势，协作编写这套丛书，它既可供有关院校当前开展继续教育选用，又能兼顾中学教师自学进修的需要，这是切合时宜的。

中学教师继续教育这项工作目前尚处于探索、研究、实践的阶段，因此，可以说这套丛书的编撰工作也同样处于探索阶段，只能随着我国继续教育事业的发展而逐步改进、完善。但编委会和编写者在调查研究和从事中学教师继续教育的基础上确定丛书的选题和内容，努力把思想政治教育放在首位，致力于学科前沿知识与基础知识同中学教育实际的结合，教育科学与心理科学同中学学科教育实际的结合，这是可取的。丛书以科学性、先进性、适用性、针对性作为努力方向，这就把教师培训与提高教育质量有机地联系起来，我相信它将会受到广大教师的欢迎。

编写《中学教师继续教育丛书》是一项开创性的工作。我们希望参编院校发挥团结协作的精神，不断实践，不断提高，共同把这套丛书编好，为中学教师继续教育事业作出贡献。

金长泽

金长泽同志系国家教委师范教育司司长

目 录

第一章 研究语言测试的必要性.....	(1)
第二章 语言测试的类别.....	(7)
第一节 由测试目的决定的测试 种类.....	(7)
第二节 由测试的方式、方法决 定的测试种类.....	(12)
第三节 由解释测试分数方法决 定的测试种类.....	(19)
第四节 语言交际能力测试.....	(21)
第三章 效度.....	(27)
第一节 内容效度.....	(27)
第二节 效标关联效度.....	(30)
第三节 构想效度和卷面效度.....	(34)
第四章 信度.....	(37)
第一节 信度的意义.....	(37)
第二节 估计信度系数的方法.....	(40)
第三节 真分数和标准测量误差...	(45)
第四节 评卷员信度.....	(47)
第五节 怎样提高测试信度.....	(49)

第六节	信度与效度的关系·····	(55)
第五章	测试结果的统计与分析·····	(56)
第一节	原始分、平均分和标准 差·····	(57)
第二节	分数分布直方图和分数 分布曲线·····	(62)
第三节	正态分布曲线和标准分 数·····	(63)
第四节	相关系数的计算方法·····	(73)
第五节	信度系数和标准测量误 差的估计·····	(80)
第六章	语音、语法和词汇知识 的测试·····	(87)
第一节	语言知识的客观测试概 述和多项选择试题的命 题要领·····	(87)
第二节	语音知识的测试·····	(99)
第三节	语法知识的测试·····	(105)
第四节	词汇知识的测试·····	(118)
第七章	总体语言能力的测试·····	(131)
第一节	完形填空和校读改错·····	(132)
第二节	听写·····	(152)
第八章	听力理解能力的测试·····	(159)
第一节	概述·····	(159)
第二节	听力理解能力测试方法·····	(162)
第九章	阅读理解能力的测试·····	(180)

第一节	概述.....	(180)
第二节	测试方法.....	(182)
第十章	书面表达能力的测试.....	(194)
第一节	概述.....	(194)
第二节	书面表达能力的测试 方法.....	(196)
第三节	书面表达能力测试的 评分.....	(199)
第十一章	口语表达能力的测试.....	(209)
第一节	概述.....	(209)
第二节	口语表达能力测试方法...	(211)
第三节	口语表达能力测试的 评分.....	(218)
第十二章	题目分析.....	(221)
第一节	测试题目的区分度和 难度.....	(223)
第二节	题目分析的步骤.....	(229)
附录 I	正态分布曲线面积和 Z 分数对照表.....	(239)
附录 II	主要术语汉英对照表.....	(250)
附录 III	主要参考书目.....	(255)

□第一章

研究语言测试的必要性

测试就是我们通常所说的考试或测验。考试和测验没有本质上的区别，都是用来测量人们掌握的知识和本领的手段。一般说，用于正式场合的测试常常被叫做考试，多半是用来在特定的人群中进行选拔或淘汰，如高等学校入学考试、毕业考试、录用人才的考试等等。用于一般场合的测试，常常称为测验，主要是为了了解人们掌握知识或技能的情况，如课堂小测验、阶段性测验、教育或教学科研工作中为获取数据而专门设计的测验等等。在教育评估中，一般不对考试和测验加以区别，把它们都看做是一种测量手段。在我们这本小册子里将使用“测试”来泛指考试或测验。有些章节，为了顺应人们称谓上的习惯，间或也使用考试和测验，但是这三者的内涵没有区别。

语言测试，顾名思义，是指测量一个人语言知识和语言能力的考试。它是应用语言学中的一个重要组成部分，其一般原则和方法普遍适用于不同语种的测试。不过这本小册子里讨论的问题和举出的实例都和我国的英语教学有关，又由于英语在我们国家是一种外语，所以，确切地说，我们这里讨论的是做为外语的英语语言测试。

语言测试和语言教学有着非常密切的关系。首先，这两者都受着特定历史阶段在语言教学界占主要地位的语言观和语言理论的制约。比如规定语言学盛行时期，人们普遍采用语法翻译法教学，语言测试就大量采用翻译、作文和语法分析的方法；而属于描写语言学理论的结构学派的产生和确立则带来了听说法和句型教学，考试上就出现了分离式客观考试；近二十年来功能学派的不断发展使人们探索交际教学法，于是语言交际能力考试也就应运而生。

其次，语言测试总会对语言教学产生或有利、或有害的巨大影响，我们把这种影响叫做考试对教学的反拨作用。测试中考什么和怎样考，会直接影响到教师的教学重点的确定，甚至影响到教材和教法的选择以及学生的学习态度和学习方法等等。大规模的公开考试往往会左右推行这一考试的整个地区的教学，带来极其深远的影响。例如国家高等学校入学英语标准化考试（MET），从1985年试行时候起，就注重测验学生运用英语的能力，扩大了检查阅读理解能力试题的量和分数比重。几年下来，全国各地英语教学普遍较以前注意培养学生的阅读能力，英语泛读教材的出版发行量大大增加，学生的阅读能力也有了较大幅度的提高。又如，由于目前某些客观条件的限制，MET还不能普遍采用口试的办法检查学生的发音和口语能力，而只是在书面试卷上考查学生的语音知识。这就有可能带来教师在涉及到听、说、读教学的语音方面重知识轻实践的结果，造成学生的语音知识和语音实践脱节，卷面上语音知识得高分，实际发音可能很差这种有害的局面。

再次，语言测试是检验语言教学、进行教学评估的重要

手段。教学方法的优劣、教材的适用性、教学大纲的合理性与科学性，无不集中地反映在教学效果，即学生经过学习之后，最终对语言知识和技能掌握的熟练程度上。客观准确地测量学生的语言知识和能力，自然就成了获取评价教学重要数据的一种方法。从语言教学的科学实验和研究方面来说，语言测试的结果，又是重要的数据来源。至于日常教学中检查学生学习成绩，决定学生的升班、留级、入学、毕业，更无一不是以测试成绩为决策依据的。所以我们完全可以说，语言教学离不开语言测试，语言测试是语言教学和语言教学科学研究的重要组成部分。

上文已经说过，语言测试是检测人们掌握某种语言知识和能力的手段。所谓语言测试，无非就是给考生规定某些特定的任务，如读一篇文章，然后回答就文章内容所提出的问题，根据要求写一篇作文，判断一个句子正确与否，把一个不完整的句子写全，或者提出问题的同时提供几个有可能做为问题答案的选择，要求考生选出自己认为是正确的答案等等。在考生完成这些任务之后，再根据事先定好的评分标准和考生答题的正确与否或答题质量的优劣，就他完成任务的情况打一个分数。这个分数就是我们对应试学生语言能力的定量化描述。如果考生 A 在一次设计质量良好的考试中得了 80 分，而考生 B 在同一次考试中得了 70 分，我们就说考生 A 的语言知识和语言能力掌握得比考生 B 要好。打个比喻来说，我们设计的一套试题，就是一把尺子。用这套试题进行测试，就是用这把尺子把所有应试学生的语言知识和能力测量了一遍。而考试分数就是测量的结果。所以研究考试的学科叫做教育测量学，在我国有人也把这门学科称为考

试学。

在明确考试是测量的手段的同时，应该强调指出的是，语言测试这样的测量，不同于对人的身高、体重、体温、跑百米的速度等等那种物理测量，而是一种智力测量，或者叫心理测量。物理测量的对象，是看得到、摸得着的物质的存在，可以进行直接测量，而心理测量的对象，是不能观察到的精神存在，只能进行间接测量。仍以上面提到的 A、B 两个学生为例。他们在一次考试中分别得了 80 分和 70 分。如果说这两个分数是对他们各自的语言水平的定量描述，那也应该看做是我们根据他们在一次特定考试中的表现，对他们的语言知识和语言能力水平进行推断而得出的结果。至于说这一推断的准确程度如何，那还需要通过我们对参加过的那次考试进行分析，并且比较他们参加过其它有关测试成绩之后才能正确地判断。测试成绩不能完全精确地反映被测试者的真正水平，这叫做测试产生了误差。凡测量都会有误差。属于间接测量的心理测量更容易产生误差，语言测试也不例外。试题本身的质量、评分方法、考场环境条件、应试人员参加考试的动机、对待考试的态度、应试策略等诸因素，都会影响考试成绩，使考试产生不同程度有时甚至是严重的误差。

经过了长时间的实践和探索，在语言测试领域已经有了一套比较系统的理论和方法，使我们有可能最大限度地减少测试中的误差，并对可能出现的误差做出接近实际的估计，从而使语言测试能够更好地为语言教学和科研服务。根据当代语言测试理论的要求，一份质量合格的语言考试试卷至少应该符合下面的五条标准：

1. 有良好的信度 信度简单地说就是考试的稳定性。用信度高的试卷多次测试同一组学生，能力强的学生总是得高分，能力差的学生总是得低分，而对不是同样水平的学生给出忽高忽低的分数。信度高的测试可以最大限度地反映学生的真正水平，它给出的分数是可以信赖的。

2. 有良好的效度 测试的效度就是测试的有效性。一份测试阅读理解能力的试卷，如果选用的阅读文章虽然文字浅显，但文章内容涉及的问题是考生所不熟悉的，甚至是难于理解的，那么这份试卷测试分数所反映的主要是学生所掌握的与文章内容有关的知识水平，而不是英语阅读理解能力。因此，做为语言测试，它的效度不高，没能有效地检测出考生的语言水平。

3. 有良好的区分能力 考试对不同语言能力水平上的学生应给出不同的分数，反映出他们能力上的差别。这就是我们所说的考试应能使考生拉开档次。对于为选拔目的而设计的测试，这一点尤为重要。

4. 对教学有良好的反拨作用 无论是教师还是学生，总是要争取良好的考试成绩，因此考试的内容和方法就会影响到他们的教或学，这一点在上文谈及测试与教学关系时已经提到。好的语言测试应该符合先进的应用语言学和教育测量学理论，从而有利于教学，把教学引到正确的方向。

5. 实践上切实可行 测试在多数情况下都不是为着一个人或少数人的。考试规模越大，牵涉的人越多，则测试需要投入的时间、人力、物力就越多。因此设计考试要考虑到可行性原则。客观条件不允许的测试，不仅会造成巨大的浪费，其信度、效度也不可能得到保证，有时甚至成了一次失

败的考试。

学习一些有关语言测试的理论，认识测试的规律，提高命题技巧，是一个从事语言教学的教师必不可少的素质修养。有了这种知识，在考试之前，如果自己设计试卷，可以帮助我们提高试题质量，减少测试中的误差，如果选用他人设计的试卷，可以使我们有一个鉴别的标准，进行科学合理的选择；在考试之后可以帮助我们科学地分析考试结果，对考试成绩做出恰当的估计，不致造成决策上的错误。这本小册子就是要在在这方面给读者提供一点帮助。在以后的各章里，我们将围绕上面提到的合格试卷的五项标准展开讨论。第二章为读者介绍语言测试的分类并说明各种不同类别的语言测试之间的区别与联系，同时着重介绍什么是语言交际能力测试。第三章和第四章分别从理论上说明测试的效度与信度、它们各自的重要意义以及估计效度、信度的方法。第五章介绍测试结果的统计与分析方法、最常用的估计效度、信度的数学公式和计算步骤，可以帮帮读者加深对一些理论概念的理解，其中一些方法和公式简单易行，也可供读者在工作中使用。第六章至第十一章是具体测试方法和评分方法的介绍。除国内一些大规模语言测试普遍采用的方法外，也有一些我们不常使用的方法。有的地方讨论了一些测试方法在理论上的依据，可以帮助读者提高使用这些方法的自觉性。最后一章讲题目分析，为读者提供改进提高试题质量的方法。书后附有正态分布曲线面积和 Z 分数对照表，供对使用标准分数感兴趣的读者使用。另外还附有本书中使用过的重要术语英汉对照表和参考书目，方便读者查阅有关资料。