

英语考试设计手册

WRITING ENGLISH LANGUAGE TEST

著者: J.B.希顿 译者: 魏进尊 林铭杰

吉林人民出版社

英语考试设计手册

J.B 希顿 著

魏进尊 林铭杰 译

吉林人民出版社

英语考试设计手册

J·B 希顿 著

魏进尊 林铭杰 译

**吉林人民出版社出版 吉林市科技进修学院发行组发行
长春市第五印刷厂印刷**

787×1092毫米32开本 11.75印张 268,000字

1986年11月第1版 1986年11月第1次印刷

印数：1—10,000册

统一书号：7091·1686 定价：1.95元

译 者 的 话

在我国,关心英语考试的人越来越多了,不仅有主持考试的部门和执考教师,还有受试的考生们。人们都想了解一点有关考试的学问,也想弄清楚外国人是怎么做的,以便借鉴。

国外较为流行的考试种类有哪些?其优缺点究竟如何?哪些应该提倡,哪些是走过的弯路?考试应当如何设计?拟题原则是什么?评分标准是什么?如何进行口试?……乃至如何建立自己的题库等等。凡此种种,都在本书讨论范围之内。所以,本书不仅适合高等和中等学校的教师们一读,也值得考生们参考。

本书作者纵览了国外大量有关英语考试的著作及资料。该书博采各家之长,既是一部综合性学术著作,又是可供随时查阅的参考手册。

在我国,专门讨论英语考试的文章和书籍极为鲜见。我们希望,奉献给读者的这本小册子,能够提供有益的借鉴,产生积极的影响。让我国的英语考试真正成为推动我国英语教学的动力吧。

本书前言至第七章前半部份为魏进尊同志所译,余为林铭杰同志所译。在翻译过程中曾得到东北师大王武君教授等的指教,亦得到有关同志的帮助,在此一并致谢。

由于译者水平所限,本书错漏之处,恳请专家与读者给予批评指正。

一九八五年十一月

前 言

长期以来，人们对待语言测试的态度都过于狭隘。直到最近，语言测试才得以正名，人们才认识到它是能够推动教学前进的具有巨大潜力的动力。整个语言考试这一课题所扣击的知识大门，的确不仅是为了更加细致地分析有关的测试法和教学法，也是为了更好地识别所教语言的性质。

这本设计英语考试的实用手册，主要是给讲课教师用的。本书不仅概述了语言测试的一般原则，还告诉教师们如何设计范围很广的测试和试题，以及如何选择最适合于他们各种目的的测试和试题。本书介绍的题型很多，并根据这些题型同所衡量的某些语言原理或技巧的关系，评估了它们的价值。要想能够获得设计恰当而又可靠的试题的能力，必须不断地实践。所以，本书后附有经过精心安排，按比例搭配的作业题，以供教师练习采用不同方法，设计各种有效的试题。希望仿照此法，能使教师更加深刻地了解把英语作为外语来测试和教学的基本原理与技巧。

术 语

测试标准和测试类型将在第九章中详细讨论。不过，读者熟悉一下本书采用的少数几个术语，则是很有用处的。

一次良好的测试应当具备的特点

有效性 (Validity) 一次好的测试，应当具备有效性：即，它应当能够衡量所要衡量的内容，而不是其它。如果一次测试能够达到这一目的，就要说它是有效的。例如，要是考生的个性能够影响执考者，还有什么有效性可言呢？

可靠性 (Reliability) 如果测试不可靠，决不可能有效。假如在第二次测试时（若是这中间教师既没有教，学生也没有学），学生的名次与前次不同，这种测试也就缺乏有效性。还有，如果两个以上的评卷教师给同一个答案评了不同的分数，这一测试也就缺乏可靠性。

区别性 (Discrimination) 许多课堂测试的主要目的，不在于区别考生，因为它的目标是把全班作为一个整体，评估其掌握某教学大纲的程度。但是，其它大多数测试（如毕业考试^① (examination) 等，都是为了区别各考生的不同能力，因而尽可能把各考生拉开档次，加以区别。

反馈作用 (Backwash, 或 Washback) “反馈”这

注：此处为区别测试 (test)，而把 examination 译作“考试”。

个词是指测试对教学的反作用。如果一次测试具有良好的反馈作用，它就能对测试之前发生的讲授和学习产生良好的影响。

测试的类型

学业进步测试 (Progress Tests) 大多数的课堂测试，都取学业进步测试形式，因为这些测试能够评价学生掌握课堂讲授内容的进步情况。学业进步测试常常用来激励学生。这些测试也使教师能够估计教学效果，帮助教师找到薄弱环节和难点。

成绩测试或成就测试 (Achievement/attainment Tests) 成绩测试或成就测试这两个术语，通常是指那些用来表明掌握某种大纲的较正式的测试（比如，学年升级测试、毕业测试、统考等）。从某种意义上讲，这些测试是以某个教学大纲为基础，衡量教学内容的，类似于课堂进步测试。但是，这些测试很少由任课教师为某一班级出题。设计这种测试的主要目的是衡量每个人的能力，而不是一种作为激励学生或巩固学习效果的测试。

水平考试^① (Proficiency Tests) 水平考试是衡量学生是否具有完成以后某项具体任务的英语水平。例如，学生的英语知识能够听懂某种用英语讲授的课程吗？或者能够胜任某项使用英语的工作吗？水平考试很少去依据学生学的是哪种教材，因为它只考虑学生将来的能力，而不考虑过去的成绩，并且这种测试常常测试许多语言学习背景大不相同的学生。

能力倾向测试 (Aptitude Tests) 能力倾向测试是

注：此处没有译成“测试”是考虑到国内某些习惯，以下同。

衡量学生开始学习一门外语时可能具备的能力，显示该生在学习一门新的语言时是否具有某种特殊的能力倾向。在辨认这种新语言的发音或者语法结构时，他是否感到很吃力呢？

诊断测试 (Diagnostic Tests) 许许多多的学业进步测试、成绩测试和水平考试，在某种程度上，都可以作为诊断测试，使教师能发现某些薄弱环节和难点，使其制订最合适的补正方案。诊断测试主要用于开始学习某种教材之前，评估学生在某些方面的知识和技能。

目 录

1	语言测试介绍	1
1.1	测试和教学	1
1.2	为什么要测试	2
1.3	应当测试什么, 按照什么标准测试?	3
1.4	测试语言技能	4
1.5	测试语言要素	5
1.6	语言技能和语言要素	8
1.7	辨认与表达	9
1.8	采样问题	11
1.9	避免捉弄学生	14
2	客观性测试	16
2.1	主观性测试和客观性测试	16
2.2	客观性测试	18
2.3	多选题: 总述	21
2.4	多选题: 题干、正确答案、迷惑答案	24
2.5	答题形式	29
3	语法和惯用法的测试	31
3.1	多选题式语法题(1): 题型	31
3.2	多选题式语法题(2): 结构辨认测试	34
3.3	设计多选题	43

3.4	设计辨认错误的多选题.....	41
3.5	设计词序题.....	50
3.6	设计填空题.....	52
3.7	设计转换题.....	56
4	测试词汇.....	61
4.1	选择试题.....	61
4.2	多选题(1).....	62
4.3	多选题(2).....	69
4.4	单词归类(有联系的词).....	72
4.5	配对题.....	73
4.6	更加客观的试题.....	79
4.7	填空题.....	81
5	听力理解测试.....	83
5.1	总述.....	83
5.2	区别音素测试.....	85
5.3	重音和语调测试.....	92
5.4	通过直观材料测试理解.....	95
5.5	叙述和对话.....	112
5.6	理解谈话和演说.....	116
6	口语表达测试.....	124
6.1	测试说话技能常见的难点.....	124
6.2	朗读.....	126
6.3	口语练习.....	129
6.4	使用图画评估口语表达能力.....	137
6.5	面试.....	141
6.6	口试的某些其它技巧.....	147
6.7	书面测试法.....	149

7	测试阅读理解	154
7.1	阅读技能的属性	154
7.2	阅读初始阶段: 配对测试	157
7.3	正误阅读测试	161
7.4	多选题: 短文	166
7.5	多选题: 较长的文章	169
7.6	填空题	181
7.7	综合填充法	185
7.8	略读	189
8	测试写作技能	191
8.1	作文测试	191
8.2	情景作文	193
8.3	作文评分	199
8.4	写作技能	205
8.5	写作的客观性测试: 关联、联接和扩展	207
8.6	写作的客观性测试: 写作规则	211
8.7	写作的客观性测试: 判断技能	215
8.8	控制性写作	224
9	测试的标准和类型	229
9.1	有效性	229
9.2	可靠性	232
9.3	区别性	235
9.4	测试的实施	236
9.5	测试要求	238
9.6	反馈作用	242
9.7	测试的种类	243
10	解释测试分数	247

10.1	频率分布	247
10.2	中间趋势衡量法	249
10.3	分散度的衡量	251
10.4	试题分析	253
10.5	猜测	261
10.6	审核	266
10.7	试题卡和试题库	267
附录		270
听写		270
翻译		272
写释义、摘要和总结		273
文学测试		274
参考书目		276
实习作业		286
本书英汉译名对照表		351

1. 语言测试介绍

1.1 测试和教学

过去，许多的测试都和教学分家，这实在是一件憾事。测试和教学两者的关系实在太密切了，二者缺一不可，缺少任何一方，工作都做不好。测试主要是作为巩固所学知识、激励学生学习的一种方法，作为方法，测试就和曾经进行过的教学结合起来了；或者作为评估学生语言能力的手段。为手段，教学则常常在很大程度上又和考试结合起来了。实际上，标准测试和统考对每个教员都有很大的影响，这些测试常常在试前促使他们决定采取何种教学方法。

考虑下面这种试题对英语教学会有怎样的影响：

Analyse into clauses, stating the kind and function of each clause.

(分析从句，说出各从句的种类和语法功能。)

We cannot go until we have finished these exercises.

同时跟下面试题的作用进行比较：

Rewrite each of the following sentences in another way but do not change the meaning. Begin each new sentence with the words given.

(用另一种方法改写句子,但不要改变意思。每个新句用所给的词开头。)

We cannot go until we finished these exercises.
When.....

前面的试题鼓励讲授“关于”语言的知识,后者则鼓励运用语言进行实践。

可以毫无疑问地说,过去的语言测试已经给语言教师带来了不良影响,由于鼓励教师讲授关于语言的知识,大大妨碍了语言学习。结果,几乎没有教师去想方设法,最大限度地向学生提供机会,进行语言本身的实践。

多亏有许多的外部考试机构,目前正在寻求衡量考生的实际语言能力,这就有时能对教学大纲和教学策略产生有益的影响。而且,不论某个公共考试团体如何关心测试对教学的影响(即反馈),这种测试的主要目的都是衡量考生使用目的语的能力。因此,这种测试是衡量和评估能力的一种主要措施。

1.2为什么要测试

前段所说测试的作用,只是对“为什么要测试”这个问题的许多解答之一。必须强调,这只是测试的许多作用之一。而且,对我们所说的讲求实践的教员来说,也许就是许多反作用之一。

尽管大多数教师也都希望评估每个人的能力,课堂测试的目的仍然与外部测试大不相同。外部测试的目的的一般只在于选拔人才而评估,课堂测试评估的目的乃是为了让教师改进教学方法,提高教学效果,让班上某些学生组或者每个学生

收效更大一些。这样的教员太多，他们总是把自己的讲授针对他错误地断定为“一般水平”的那些学生，而不考虑处于两极的学生们能力如何。

一次良好的课堂测试也有利于准确地确定全班或个别学生感到困难的地方。就象是医生必须诊断其病人的疾病一样，教师也必须对他的学生的薄弱环节和难点进行“确诊”。如果教师不能辨别、分析学生在使用目的语中出现的错误，他就完全不可能恰如其分地预见、补课及增加练习以提供任何帮助。

测试也应当能使教师发现，在执行某个语言教学计划时班上的难点。这样，教师就能够估计教学大纲以及他目前使用的教学法和教材的效果。例如，测试结果可能表明，这种语言教学大纲的某些地方没能充分地考虑到语言学习者的困难，或者由于某种原因没有突出出来。

应当尽可能把测试用于衡量学生的能力，无论如何都不应捉弄学生，这就能够有效地激励学生。一次设计得很好的课堂测验，能让学生有机会显示他的能力，认识和运用正确的语言形式。如果测试之后能尽快地给出学生测试结果的详细情况，他们也就能从错误中学到东西，巩固所学句型。这样，一次良好的测试就可以作为一种很有价值的教学手段。

1.3 应当测试什么，按照什么标准测试？

现代语言理论的发展，已经使语言教师和测试人员认识到分析所测语言的重要性。当代的描述语法（尽管主要还不是以外语教学为目的）正在取代古老的，以拉丁语为基础的法规式语法；语言学家们正在研究语言技巧和语言行为模式

的整个复杂系统。的确，语言技能太复杂了，同其它许多非语言技能（手势、眼睛的运动等）的关系太密切了，似乎进行任何评价都不可能将它们割裂开来。一个人在某一特定时间、特定条件下，总是要讲话，要进行思想交流的。没有这个上下文，语言就可能失掉其本身的许多特殊含义。

在设计测试之前，重要的问题就是采用什么标准。学习外语的学生应当按什么标准要求呢？例如，应当要求学习外语的学生象运用母语那样自然而流畅地进行交际吗？学习第二语言的人，有些习惯在其母语中完全正确，同是这些习惯，在第二语言中是不是错误的？到底什么是“地道”的英语呢？

从前在书面语言测试中，甚至给英语为母语的人搞了一些不切实际的标准，常常要求具备英语大作家和大批评家那样纯熟的技能。在模拟第一语言书面英语测试时，他们盼望外国学生具有的能力很不现实，曾经要求学生们用自己的话改写某些文学经典著作，也曾要求用超出他们能力的语言写作具有独创性的论说文。

1.4 测试语言技能

语言交际的四大技能，一般规定为：听、说、读、写。在许多教学环境中，最理想的是：在训练读写之前，先进行广泛的口语训练。在这种情况下，重要的是让试题编写者运用出现在讲说这种语言中出现的问题类型，即，测试学生驾驭口语中碰到的句子结构的能力，和对某个给定的刺激做出正确反应的能力。

参加传统测试取得成功，常常只不过表明考生能够很好

地完成他所参加的测试答题，而几乎不意味着其它。例如，编写段落大意的练习（这是许多重要测试的主要试题之一），所衡量的技能跟测试与答题技巧的密切性，超过了同现实生活环境中所用语言的关系。从这个意义看，传统的测试几乎不可能告诉我们关于学生掌握所学语言的流畅程度和运用能力，虽然它也能够标志着作为一个学生所需要的某些技能。

评价这四项技能，可以采取的测试形式有：

（1）听力（听觉）理解。向考生提供单个发音、对话、谈话及讲演等；

（2）讲话能力。常常采用面试，看图说话，以及朗读等形式；

（3）阅读理解。提出问题，测试学生对某篇文章的理解；

（4）写作能力。常常是写作论说文、信函和报告等。

试题设计人员的任务就是评估在不同水平阶段，这些技能的相对重要性，提供一种精确的手段来衡量学生在这些技能的发展过程中所取得的成就。许多拟题者都认为，只要能对各种技能分别衡量，就能够圆满地达到目的。但是，要想把这些技能互相分开，常常是编其困难的，因为这四种技能的划分是人为的，而且这个概念过于简单化，理由是它本身只包含交际中涉及到的问题。

1.5 测试语言要素

在教学和测试中，为了把语言技能的构成成份分开，我们可以仔细看看下面这些要素：