

An Introduction to English Language Testing

简明英语测试教程

邹申 杨任明 著



高等教育出版社



An Introduction to English Language Testing

简明英语测试教程

邹申 杨任明 著

高等教育出版社

内 容 提 要

《简明英语测试教程》为高等师范院校英语或英语教育专业教材，也可供高校或中学英语师资培训使用。

本教程在介绍国内外语言测试最新理论研究及实践的基础上，结合我国英语教学实际及学生特点，着重讲解语言测试流程中的诸多环节。全书共13个章节，书后附有英汉对照测试术语表和测试公式表，可供一个学期使用。每章包括思考题、英文摘要、正文和练习四个部分。思考题起到课前启发引导的作用，英文摘要简要地介绍该章节内容及一些测试术语。正文是章节的主要部分，讲解一个主题。练习则是让学生通过实际操作掌握或巩固每章节讲述的要点。每章节的练习既有相对的独立性，又与全书其它部分的练习构成一个有机整体。

图书在版编目(CIP)数据

简明英语测试教程 / 邹申, 杨任明著. —北京: 高等
教育出版社, 2000. 7

ISBN 7-04-008922-X

I. 简... II. ①邹... ②杨... III. 英语-测试-师范
大学-教材 IV. H319

中国版本图书馆 CIP 数据核字 (2000) 第 63774 号

简明英语测试教程

邹 申 杨任明 著

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮 政 编 码 100009

电 话 010-64054588

传 真 010-64014048

网 址 <http://www.hep.edu.cn>

经 销 新华书店北京发行所

印 刷 高等教育出版社印刷厂

开 本 850×1168 1/16

版 次 2000 年 7 月第 1 版

印 张 12.75

印 次 2000 年 7 月第 1 次印刷

字 数 346 000

定 价 20.00 元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换。

版 权 所 有 侵 权 必 究

使 用 说 明

本教程为高等师范院校英语或英语教育专业教材，也可供高校或中学英语师资培训使用。

本教程在介绍国内外语言测试领域最新理论研究及实践的基础上，结合我国英语教学实际及学生特点，着重讲解语言测试流程中的诸多环节，如考试的宏观、微观功能，考试总体设计，单项语言能力/技能的测试、命题、施考，考试分析，考试信息反馈。

本书共13章，书后附有英汉对照测试术语表和测试公式表，可供一个学期使用。每章包括思考题、英文摘要、正文和练习四个部分。思考题起到课前启发引导的作用，英文摘要简要地介绍该章节内容及一些测试术语。正文是章节的主要部分，讲解一个主题。讲解力求做到通俗易懂，简明扼要，对实际操作具有指导性。练习部分的目的是让学生通过实际操作掌握或巩固章节内讲述的要点。每章节内的练习既有相对的独立性，又与全书其它部分的练习构成一个有机整体。

本教程的编写宗旨是通过对考试流程中主要环节的介绍和演示，让使用者了解并掌握语言测试的基本原理及具体操作方法，以便使他们在未来教学实践中提高命题水平和考试质量。因此，我们建议在教材使用中要体现以人为本的思路，采取启发式教育方法，鼓励学生积极思考，在学习过程中发现和解决问题。课堂教学可以采取讨论形式，让学生充分发表自己的见解。课后练习也不一定拘泥于某一固定模式，可以采取先集体讨论，随后个人或小组完成的形式。总之，我们希望通过本教程的学习，学生不仅了解一些测试理论概念，而且初步具备测试的实际操作能力。

编 者

2000年5月

目 录

第一章 英语测试概述：过去、现在与未来	1
1.1. 四种英语语言测试法	2
1.2. 近20年的发展动态	4
1.2.1. 考试方式的影响	4
1.2.2. 考试分析	4
1.2.3. 考生特征	5
1.2.4. 语言能力性质的再认识	5
1.3. 当代国内外大规模考试一览	6
1.3.1. 全国公共英语等级考试体系	6
1.3.2. 大学英语四、六级考试	10
1.3.3. 英语专业四、八级考试	12
1.3.4. 美国托福考试	14
1.3.5. 英语雅思考试	14
1.3.6. 语言测试前景展望	15
练习	17
第二章 考试功能及其类别	18
2.1. 测量、考试与评估	19
2.2. 考试功能	20
2.2.1. 教学功能	20
2.2.2. 科研功能	22
2.3. 考试类别	22
2.3.1. 考试目的	23
2.3.2. 考试方式	24
2.3.3. 语言测量形式	25
2.3.4. 评分方式	26
2.3.5. 考分解释	26
练习	28
第三章 考试要素	31
3.1. 考试信度	32

3.1.1. 重复测试法	34
3.1.2. 平行卷测试法	35
3.1.3. 对半分析法	35
3.2. 考试效度	36
3.2.1. 表面效度	36
3.2.2. 内容效度	36
3.2.3. 尺度关联效度	37
3.2.4. 结构效度	38
3.2.5. 效度与信度的关系	39
3.3. 真实性	39
3.4. 交互性	40
3.5. 后效作用	41
3.6. 可操作性	41
练习	42

第四章 考试规范的制订	47
4.1. 什么是考试规范?	48
4.2. 为什么需要考试规范?	48
4.2.1. 考试规范与考试效度	48
4.2.2. 考试规范与总体设计	48
4.2.3. 考试规范与考试评估	49
4.2.4. 考试规范与考试稳定性	49
4.3. 考试规范包括哪些内容?	49
4.4. 如何制订考试规范?	52
4.4.1. 考试目的	52
4.4.2. 考试能力范畴	53
4.4.3. 考试对象	53
4.4.4. 考试内容	54
4.4.5. 考试方式	54
4.4.6. 考题形式	54
4.4.7. 试卷结构与分值	56
4.4.8. 考试时间	56
4.4.9. 评分标准	57

4.4.10. 考试实施	57
4.4.11. 考试命题细目表	57
练习	59
第五章 测试评估手段及其功能	61
5.1. 选择答题类评估方法	62
5.1.1. 正误判断题	62
5.1.2. 匹配题	64
5.1.3. 多项选择题	66
5.2. 非选择答题类评估方法	69
5.2.1. 填空题	69
5.2.2. 转换题	72
5.2.3. 简答题	74
5.2.4. 论述题	75
5.3. 个性化答题类评估方法	75
5.3.1. 个别检查	76
5.3.2. 自评与互评	76
5.3.3. 学期作品评估	77
练习	79
第六章 阅读测试	81
6.1. 什么是阅读?	82
6.1.1. 阅读目的	82
6.1.2. 阅读方式	83
6.1.3. 阅读材料	83
6.2. 什么是阅读能力?	83
6.3. 如何编制阅读试卷?	85
6.3.1. 考试规范	85
6.3.2. 考试题型	87
练习	92
第七章 写作测试	96
7.1. 什么是写作能力?	97

7.2. 如何编制写作试卷？	98
7.2.1. 考试规范	98
7.2.2. 考试题型	101
7.3. 如何评阅作文？	107
7.3.1. 评分标准	107
7.3.2. 评分信度	110
练习	111
第八章 听力测试	114
8.1. 听力活动特点	115
8.1.1. 听力信息特点	115
8.1.2. 什么是听力理解？	115
8.2. 什么是听力？	116
8.3. 如何编制听力试卷	117
8.3.1. 考试规范	117
8.3.2. 考试题型	119
8.4. 听力命题有关事宜	126
8.4.1. 选材与录音	126
8.4.2. 考点确定	126
练习	127
第九章 口语测试	129
9.1. 什么是口语能力？	130
9.1.1. 口语活动的特点	130
9.1.2. 口语能力的构成	130
9.2. 如何编制口语试卷？	131
9.2.1. 考试规范	131
9.2.2. 考试题型	133
9.3. 如何评价口试？	136
9.3.1. 评分标准	136
9.3.2. 评分信度	138
练习	139

第十章 试卷设计与施考事项	143
10.1. 试卷设计	144
10.1.1. 考试项目顺序	144
10.1.2. 项目提示语	145
10.1.3. 卷面设计	146
10.2. 施考事项	149
10.2.1. 制订施考细则	149
10.2.2. 培训监 / 施考人员	149
练习	150
第十一章 考试成绩反馈	155
11.1. 考试成绩	156
11.1.1. 原始分数	156
11.1.2. 转换分数	156
11.1.3. 成绩报告形式	157
11.2. 及格标准	158
练习	159
第十二章 考试成绩分析(1)	161
12.1. 分数的频数分布	162
12.2. 分数的集中量	163
12.2.1. 算术平均数	164
12.2.2. 中位数	164
12.2.3. 众数	165
12.2.4. 算术平均数、中位数、众数三者之间的关系	165
12.3. 分数的差异量	166
12.3.1. 全距	166
12.3.2. 标准差	166
12.3.3. 差异系数	167
练习	168
第十三章 考试成绩分析(2)	170
13.1. 题目难易度	171

13.2. 题目区分度	172
13.3. 选择项分析	174
练习	177
测试术语表	181
测试公式表	186
参考书目	190

第一章 英语测试概述：过去、现在与未来



思考题

1. 英语测试领域里是否有不同的语言测试法？
2. 在过去20年中英语语言测试界所关注的热点有哪些？
3. 当今国内外有哪些著名的考试？它们各自有什么特点？

Abstract in English

This chapter offers a brief survey of English language testing in the following three aspects: general approaches to English language testing, major areas of concern in research and practice in the past twenty years or so, and main features of some of the current large-scale English language tests. The purpose of the chapter is to help you form a clear picture of what has happened and what is happening in the field of English language testing.

测试是英语语言教学中再熟悉不过的一种评估手段：教师教完一课后需要进行阶段小测验，学期进行至一半时要进行期中考试，学期结束时则要举行期末大考。有的考试只要求学生在所给的选择项中选出最佳或正确选项，有的则要求学生写出答案。在出题的时候我们或许会问自己：为什么选择某一考试形式？或者当今英语测试所关注的焦点是什么？英语测试近年来的发展动态及其趋势如何？在讲解英语测试的有关问题之前，我们认为有必要先回顾一下英语测试领域中存在的诸多语言测试观，过去20年中英语语言测试界所关注的热点，并在此基础上介绍当今国内外一些著名的考试及其特点，这对我们了解英语测试，在教学中正确、恰当地使用测试这个手段是大有裨益的。



1.1 四种英语语言测试法

第一种是写作—翻译法 (the essay-translation approach)。这种方法主要有以下特征 (Heaton, 1991: 15): 1) 对测试的技能或专长没有专门要求，主要是依靠教师的主观判断力；2) 试卷通常包括翻译、写作和语法分析等项目；3) 试卷内容带有较浓厚的文学或文化色彩；4) 试题一般采用书面回答形式，试卷需人工评阅。由于写作—翻译法不强调测试的科学性或理论性，故有时被一些学者称为语言测试的前科学阶段。

第二种是结构主义/心理测量法 (the structuralist-psychometric approach)。这种测试法以结构主义语言学为其理论基础，强调分别测试不同的语言成分，比如语音、语法和词汇都可以脱离上下文进行单独测试。此外，听说读写等项语言技能也可以分开测试，因为这个测试法的重要特征是一道题可以单独测试一个语言成分或技能。这种方法的另一大特点是采纳了心理测量学的一些方法，强调语言测量的可靠性和客观性。结构主义 / 心理测量法在语言测试中的典型表现形式是多项选择题，一种既能达到一题测试一成分的要求，同时又适合于进行考后统计分析的题型。

第三种是综合法 (the integrative approach)。综合法有如下特点 (Heaton, 1991: 16): 1) 语言测试要在一定上下文 (context) 中进行；2) 不在测试中刻意追求区分各单项语言技能或能力，而是强调两项或两项以上语言技能的综合评估。较能体现综合法的考试题型有完型填空、听写、翻译、写作等。以完型填空为例，这类形式可以同时测试学生的语法和词汇知识以及阅读理解能力。此外，语法和词汇知识以及阅读理解能力的测试在一个特定的上下文（即所给的完型填空材料）中进行。这里需要指出的一点是，上下文环境在综合法中不是真正意义上的语用环境，只是起到辅助测试语言知识的作用。

第四种是交际法 (the communicative approach)。有学者 (Heaton, 1991: 19) 认为交际法与综合法在某种程度上有相似之处，即两者都强调语言的意义而不是语言的形式和结构，但同时两者之间又存在根本区别：交际法更注重语言在交际过程中的使用。这里所牵涉到的两个基本概念是：usage和use。根据 Carroll 的观点 (1983: 7)，usage主要指语言形式和结构，而use则是有关语言的交际功能以及使用。语言运用是语言学习的最终目标，语言形式的掌握是实现这一目标的方法。因此，衡量一个人语言熟练程度的最终标准是看此人能否在语言使用环境中有效地进行交际。交际法包括usage成分，但更多的是评价学生在特定语言使用环境中的交际能力 (use)。

随着交际教学法在语言教学领域里日渐盛行，从20世纪80年代起语言测试界也逐步受到其影响，并出现了一系列旨在测试交际能力的探索性考试 (Bachman, 1991: 677)。纵观这类考试，大致有以下几个特点 (Bachman, 1991: 678; Heaton, 1991: 20-21; Carroll, 1983: 8): 1) 考试内设计“信息沟” (information gap)，要求学生通过各种已嵌入的信息来获取未知信息。2) 考试任务或项目的关联性 (task dependency)，

即一个项目的完成要建立在前一项目的基础上，比如，学生先听一个电话交谈，然后根据交谈的内容写一封信。3) 考试强调针对性，根据学生的具体需求设计考试内容，比如在专门用途英语 (English for Special Purposes) 考试中，要根据学生的具体语言需求，确定测试内容、所需语言交际模式及其相关语言技能。4) 侧重更广泛地测试语言能力，包括语言知识、语言功能、语言使用的合适性等等。5) 考试采用定性评估方式 (qualitative modes of assessment)，以取代定量评估方式 (quantitative modes of assessment) 或作为其补充。以大学英语口语考试的标准描述为例 (大学英语口语考试大纲及样题，1999: 4)：

	语言准确性和范围	话语的长短和连贯性	语言灵活性和适切性
5分	<ul style="list-style-type: none"> ● 语言和词汇基本正确； ● 表达过程中词汇丰富、语法结构较为复杂； ● 发音较好，但允许有一些不影响理解的母语口音。 	<ul style="list-style-type: none"> ● 在讨论有关话题时能进行较长时间的、语言连贯的发言，但允许由于无法找到合适的词语而造成的偶尔停顿。 	<ul style="list-style-type: none"> ● 能够自然、积极地参与讨论； ● 语言的使用总体上能与语境、功能和目的相适应。
4分	<ul style="list-style-type: none"> ● 语法和词汇有一些错误，但未严重影响交际； ● 表达过程中词汇较丰富； ● 发音尚可。 	<ul style="list-style-type: none"> ● 能进行较连贯的发言，但多数发言较简短。 ● 组织思想和搜寻词语时频繁出现停顿，有时会影响交际。 	<ul style="list-style-type: none"> ● 能够较积极地参与讨论，但有时内容不切题或未能与小组成员直接交流； ● 语言的使用基本上能与语境、功能和目的相适应。
3分	<ul style="list-style-type: none"> ● 语法和词汇有错误，且有时会影响交际； ● 表达过程中词汇不丰富，语法结构比较简单； ● 发音有缺陷，有时会影响交际。 	<ul style="list-style-type: none"> ● 发言简短； ● 组织思想和搜寻词语时频繁出现较长时间的停顿，影响交际，但能够基本完成交际任务。 	<ul style="list-style-type: none"> ● 不能积极参与讨论，有时无法适应新话题或讨论内容的改变。
2分	<ul style="list-style-type: none"> ● 语法和词汇有较多错误，以致妨碍理解； ● 表达过程中因缺乏词汇和语法结构而影响交际； ● 发音较差，以致交际时常中断。 	<ul style="list-style-type: none"> ● 发言简短且毫无连贯性，几乎无法进行交际。 	<ul style="list-style-type: none"> ● 不能参与小组讨论。

标准对每一分数段的三个方面做了详细的文字描述，这有助于教师和学生了解实际能力水准和问题的所在。另外在定性评估方式中，对某一学生能力的评估主要是看他本人达到标准的程度，而不是通过与同组内其他同学的比较。

定性评估方式属于尺度参照性质 (criterion-referenced)，定量评估方式属于常模参照性质 (norm-referenced)，然而交际法并不排除后者在测试中的作用。Carroll (1983: 10)认为，确定具体的交际行为标准属于尺度参照性质，但在测试项目分析的某些方面和考试成绩的解释上我们仍需要采用属于常模参照性质的统计方法。



1.2. 近20年的发展动态

自80年代以来英语语言测试进入了一个活跃发展时期，语言测试逐渐得到人们的重视和关注，有关研究成果不断涌现，其中的新思路、新方法对进一步推动英语语言测试研究起到了积极作用。本小节将简要地介绍近20年来的一些发展动态。

1.2.1. 考试方式的影响

近年来的一个研究焦点是考试方式方法对考试成绩的影响，一些学者在这方面开展了一系列的研究(Bachman & Palmer, 1981:67-86; 1982: 449-465; 1988: 260-275; Clifford, 1981: 62-70; Shohamy, 1983: 527-540; 1984: 147-170)。他们的研究成果表明考试方式方法对成绩和被试能力都会产生影响。在开展第一手研究的同时，这方面的理论研究也有所突破：Bachman (1990: 111-159)在他的著作中提出了一个考试方式影响面的理论框架。理论框架包括五类因素：考试环境 (the testing environment), 考试指导语 (test rubric), 语言输入性质 (the nature of the input), 预期回答性质 (the nature of the expected response)以及输入与回答的关系 (the relationship between input and response)。

1.2.2. 考试分析

过去20年里在考试统计分析方面也已有长足的发展，包括各种定性和定量的考试分析手段。定量分析手段有项目反应论 (item response theory), 因素/差异归纳论 (generalizability theory), 以及尺度参照测量法等 (criterion-referenced measurement) (Bachman, 1991: 676)。定性分析手段方面的典型例子是个人自述法 (introspection)，一种通过考生自我描述来了解应试策略和答题过程的方法。这种方法可以帮助人们从另一侧面或层次上，了解到定量分析手段尚无法提供的信息，故已引起许多语言测试专家的兴趣，并在研究中运用这种方法。现有研究表明应试者采用不同的策略来完成答题，且策略的使用直接关系到成绩好坏 (Alderson, 1988; Anderson et al, 1991: 41-66; Cohen, 1984: 70-81; Nevo, 1989: 199-215)。已有学者出版专著，介绍在语言教学与测试研究中的一些定性分析方法，

比如面试、问卷调查等，并且附有实例 (Weir & Roberts, 1994: 140-163)。

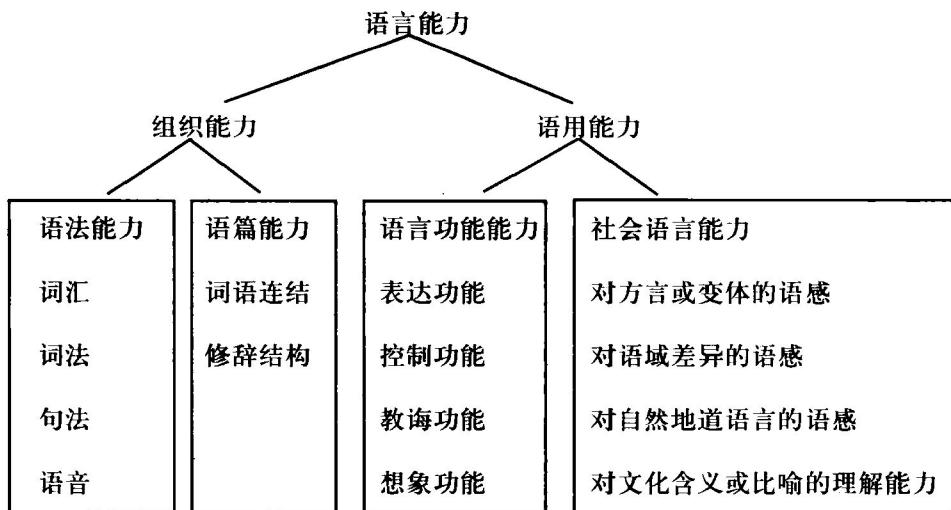
1.2.3. 考生特征

考生特征也在过去20年内成为语言测试工作者所关注的一个热点，是众多研究的对象。其中有些研究表明，文化、语言和种族背景不同会导致考试成绩上的差异 (Alderman & Holland 1981; Chen & Henning, 1985; Politzer McGroarty, 1985; Swinton & Powers, 1980)。因此，考生特征已成为考试设计过程中的一个考虑因素。Bachman 和 Palmer (1996: 111) 把与考试设计有关的考生特征归纳为以下四类：

- 1) 考生个人特征（年龄、性别、母语、受教育程度、参加某一考试的经历或准备程度等）；
- 2) 背景知识；
- 3) 语言能力；
- 4) 对考试态度（答题时的积极性或合作程度等）。

1.2.4. 语言能力性质的再认识

对语言能力性质的重新认识，也是过去20年中语言测试领域里讨论和研究的重点。80年代初Oller (1979)提出了“能力单一性的假设” (the unitary competence hypothesis)，认为语言能力是一种单一性的整体能力。Oller的假设曾一度被人们所接受，但后来受到实际研究的质疑和挑战。最后，连Oller本人也放弃了这种观点 (Bachman, 1991: 673)。现在通过实际研究和理论探讨，被人们普遍接受的观点是：语言能力是多成分的 (multicomponential)，由一系列相互关联的具体能力和一总体能力组成 (Bachman, 1991: 673)。在实际研究方面有许多实例，比如Buck (1990)的研究证明，在满足一定条件的情况下，阅读和听力能力可以区分开来；高等学校英语专业考试 (TEM) 效度研究也从不同的方面显示语言能力的多成分性 (邹申等, 1997)。在理论方面，Bachman (1990: 81-109) 提出的交际语言能力模式 (communicative language ability) 博采众长，全面地描述了语言能力的性质及其组成部分。这个模式可以说是迄今为止最完善的一个理论模式，它代表了90年代语言测试界在探讨语言能力性质方面的进展。从根本上说，Bachman的模式建立在语言能力可分隔性假设 (the divisibility hypothesis)之上。他的模式包括三个组成部分：语言能力 (language competence)，策略能力 (strategic competence) 和心理生理机制 (psychophysiological mechanism)。每一部分又可再进一步划分，比如语言能力可分成以下几个方面：



策略能力包括判断能力、计划能力和实施能力；心理生理机制则指在语言使用中的神经和生理过程，比如区分视觉和听觉途径，或产出和接受方式等。

目前，Bachman的语言能力理论模式已成为语言能力测试和考试效度评估的理论依据或参照框架之一 (Bachman & Palmer, 1996; Bachman et al, 1995)。



1.3. 当代国内外大规模考试一览

1.3.1. 全国公共英语等级考试体系

全国公共英语等级考试体系（Public English Test System, 以下简称PETS）是由教育部考试中心在1997年以来设计、开发的系列考试。

PETS考试的目标和宗旨简要归纳如下（PETS考试大纲, 1999: 4）：

1) 建立一个覆盖不同层次需要的，拥有五个级别的目标参照性英语考试体系。

2) 完善各种考查手段，尤其是听、说方面的考查方法，确保各个级别都能考查考生听、说、读、写的能力。

3) 根据我国英语教学实际和社会发展的需要，在同一能力量表的基础上合理设置各级考试的评价标准。

4) 采取有效技术手段，确保相同级别不同考次之间考生成绩的等值；逐步将有关的升学、出国、自学等考试联系起来，使考生成绩使用达到多样化。

PETS考试共分五个等级：一级相当于初中毕业时的英语水平；二级相当于普通高中优秀毕业生的英语水平；三级相当于大学公共英语4级水平；四级相当于大学公共英语6级或6级以上水平；五级相当于大学英语专业4级的水平。PETS每个等级均由笔试和口试两部分组成，以下为每一级口、笔试的结构表：

PETS一级笔试

部分	节	题型	题量	权重(%)	时间(分钟)
I. 听力	A	多项选择题 (三选一)	10	30	20
	B	多项选择题 (三选一)	15		
II. 英语知识运用	A	多项选择题 (三选一)	15	25	20
	B	完型填空 多项选择题 (三选一)	10		
III. 阅读理解	A	搭配题	5	30	30
	B	R/W/DS (正/误/未提及)	5		
	C	多项选择题 (三选一)	10		
IV. 书面表达	A	改写句子	3	15	20
	B	写简短信函或 便条	1		
总计			73+1	100	90

PETS一级口试

节	时间(分钟)	形式	分数
A	3	口试教师与考生对话	5
B	5	两考生对话	

PETS二级笔试

部分	节	题型	题量	权重(%)	时间(分钟)
I. 听力	A	多项选择题 (三选一)	5	30	20
	B	多项选择题 (三选一)	15		