

大学英语测试分析报告

栗芹芹

(郑州西亚斯学院外语学院, 河南 郑州 450000)

摘要: 大学英语在国家的素质教育及对外交流中所起的作用无法估量, 显然, 大学英语的教学在此过程中显得尤为重要。在英语教学改革的过程中, 语言测试始终是重要又必要的一环, 同时也在不断地发展和完善。本研究基于语言测试学理论, 对本人所带班级 72 名学生的期中测试进行分析, 借助语言测试分析方法和 SPSS13.0 统计分析软件, 主要从试卷结构、试卷质量检验指标、试卷成绩、及对教学的总结和反思几个方面进行逐一分析, 以便能从中发现一些问题, 希望可以促进大学英语教学经验的积累和教学质量的提高, 从而更好地帮助学生提升英语综合水平。

关键词: 大学英语; 语言测试; 分析

改革开放以来, 我国对外交流的需求日益增长, 因此, 国家很多政策发生了改变, 在人才培养方面, 对英语的综合能力也有了更加具体的说明和要求, 基于此, 我国英语教育不断做出努力和改革, 培养出了很多精通英语的人才, 基本满足了英语对外交流的需求, 而且还提升了国人的整体素质, 取得了举世瞩目的成就, 促进了我国经济社会的发展。

语言测试是检验外语教学和学习成果的重要方式, 在外语的教和学中起着非常重要的作用。在外语教育不断改革的过程中, 语言测试也在不断改革、发展和完善。语言测试是大学英语教育质量的衡量标尺, 能够整体上反映出学生的学习水平, 是大学英语改革的重要参考因素。通过大学英语测试, 不仅能够了解学生对知识点的学习情况, 还能了解教师教的情况。例如, 它能够检验学生是否学会了已经学过的知识, 从而帮助学生发现学习中不容易发现的问题, 帮助学生高效提升英语的综合应用水平。它还能够帮助教师检验教学的效果, 得到教学的反馈, 帮助教师总结教学的经验, 探索教学过程中的规律性, 从而完善教学方法, 提升教学质量, 从而引导学生更好地学习。

语言测试的制定基于国家规定的教学大纲和教材, 因此, 语言测试的不断发展和改革定能促进教材和课程设计的科学化和规范化, 从而促进学生的科学发展和培养。测试设计者基于科学的语言测试理论和国家明确的人才培养规划和目的, 会相应地设计合适的题目和对应的难度, 来考察学生的各项能力, 从而促进教材和课程设置的科学性。

本文基于语言测试学的理论, 借用刘润清提到的语言测试统计分析方法(刘润清, 1999), 使用 SPSS13.0 统计分析软件, 从试卷结构、试卷质量检验指标分析、试卷成绩分析、对教学的总结和反思四个方面对本人所教学生的大学英语测试进行具体详细的研究和分析。

本研究测试的内容是非英语专业大学英语(一)期中测试, 测试的对象是非英语专业的大学一年级本科生, 参与测试人数共有 72 人(除去未参加考试的学生)。

一、试卷结构

本研究当中所用的英语测试试卷包括听力部分和笔试部分, 共包含七个题型, 总分 100 分。在第一大部分中, 听力部分(25 分)的题型包括选择题和填空题。其中, 选择题的类型既有听短文又有听对话的形式, 填空题的类型包括听短文, 共占比 25%。在第二大一部分中, 笔试部分(75 分)的题型分为阅读理解(10%)、单向选择(10%)、深度阅读(20%)、翻译(10%)和作文(10%), 共占比 75%。另外, 客观题部分是试卷中的第一大题至第五大题,

主观题是试卷中的第六、七大题。

二、试卷质量检验指标分析

测试学理论指出, 难度、区分度、信度和效度可以作为语言测试的主要检验指标来衡量试卷的质量和水平。因此, 通过研究、探索和检验本套试题有关这几个方面的指标, 可以直接判断出该测试题目是否符合该学段学生。

(一) 难度

难度指的是试卷上题目的难易程度。在测试学中, 一般情况下, 试卷难度指的是回答试题正确的人数与参加该考试的总人数之比。用公式 $P=R/N$ 来表达, 其中, 字母 P 代表试卷中试题的难度; R 代表答对该试题的人数; N 代表的是参加该测验的总人数。然而, 这个公式不是适合所有的题目, 它常用于客观试题。对于主观性试题, 则采用公式 $P=A/B$, 其中 A 为该题得分的平均分, B 为该题的满分值。

测试学中, 可以用难度系数来表示题目的难易程度。难度系数越大, 表明该题越简单, 反之则越难。研究发现, 题目的平均难度一般在 0.5 左右浮动, 最好控制在 0.3-0.7 之间才算合理。难度系数低于 0.3 或者高于 0.7 的题目过多的话, 则不符合测试的要求(刘润清, 1999)。因为难度值能够直接表明试题的难易程度, 难度系数越是接近 0, 就表明测验的难度越大, 能够做对题目的人数就越少, 测验的最终得分就越是集中在低分区域, 那么其分数分布就不会呈现正态分布; 相反的情况下, 难度系数的值越无限靠近 1, 则代表试题的难度越小, 难度小的话, 能够正确回答试题的人数就会明显增多, 最后测验的分数就会集中在高分区域, 分数分布就越呈现偏负态分布(王汉澜, 1987)。

在英语测验中, 选择题是一种主要的题型, 不管是语法填空、仔细阅读、快速阅读、听力理解还是完形填空, 都设置成了选择题的形式。英语测验中的选择题一般包括一个题干和多个选项共同组成, 在多个选项中, 其中一个为最佳答案, 其余的选项我们可以称之为起干扰作用的选项。多项选择题目一方面考查学生对知识的掌握情况, 另一方面还考查学生在面临多种选择并且受到周围因素的干扰下, 能否做出正确的判断。

下面先以本次试卷的多项选择题为例, 计算客观性试题的难度:

根据对答题卡中第 37 题的原始数据统计。参与本次测验的总人数为 72 人, 答对本题的人数为 49 人。则 $R=49$, $N=72$, 将其代入公式 $P=R/N$ 中, 则 $P=0.68$ 。所以本试题的难度是适中的, 并且题目是有效的。

根据对答题卡中第 46 题中的原始数据统计, 参与本次测验的总人数为 72 人, 答对本题的人数为 31 人。则 $R=31$, $N=72$, 将其代入公式 $P=R/N$ 中, 则 $P=0.43$ 。所以本题的难度是适中的, 并且

题目是有效的。

下面再以本次试卷第 76 主观题（带上雨衣，以防下雨）为例，计算其难度：

根据对答题卡的原始统计，本题的平均得分为 0.55，本题的满分为 1.5。则 $A=0.55$ ， $B=1.55$ 。将其代入公式 $P=A/B$ 中，则 $P=0.37$ 。所以本题的难度适中，并且题目有效。

（二）区分度

在测量学中，有很多概念，区分度是其中很重要的一个，它是衡量题目质量的主要指标之一。如果某一套试题区分度高的话，那么就可以明显地把不同水平的考生区分开。因此，重要的考试

试卷都要有达到要求的区分度才算合格。一般情况下，区分度的范围为 $[-1, 1]$ 。在普通考试中，区分度的数值不能比 0.2 小，但对于某些非常重要的考试而言，区分度不能小于 0.3，高于 0.4 的话最佳。客观性试题的区分度计算公式为：在语言测试中，区分度符号用字母 D 来表示，高分组通过测试的人数之比用字母 PH 来表示，低分组通过测试的人数之比用 PL 来表示，所以，一般用公式 $D=P_H-P_L$ 来计算客观性试题的区分度。统计结果显示：高、低分组的人数在总人数中各占 27%。以本次试卷中第 47 客观性试题为例，计算其区分度：发现参与本次测验的人数为 72 人，则高、低分组人数各为 20 人。

表 1. 本次检测人数中高低分组情况

高分组	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	B	C
低分组	B	A	B	B	B	C	C	B	B	B	A	C	B	B	B	B	A	C	D

$P_H=19/20=0.95$ $P_L=12/20=0.6$ 将其代入公式 $D=P_H-P_L$ 中，则 $D=0.35$

所以，此题的区分度为 $0.3 < D=0.35 < 0.4$ 。本套测试题为期中考试的试题，属于比较重要的考试，该题的区分度 $D=0.35$ ，大于 0.3，说明此题的区间度良好，但是该题的区分度小于 0.4，因此也是需要进一步改进和提升的。

（三）信度

信度是评估语言测试的重要指标之一。信度一般指可靠的程度。语言测试的信度则是指测验结果的可靠程度。在语言测试中，可以采用内在信度检验中的分半法对测验试卷信度进行检验。内在信度检验中的分半法指的是把一个试卷的题目番号分为奇数题和偶数题。计算时分别得到一个奇数题总分和一个偶数题总分，然后求所有参加考试者奇数题总分与偶数题总分的相关系数。最后使用 Spearman-Brown 公式进行校正，求得整个测验的信度系数。信度系数在 0 和 1 之间，系数越高表明试题的内在一致性就越强，测试的结果就越可靠。

本次测验中 72 名学生的考试分数已在附件中列出，这其中包括每一个人的奇数题总分（X）和偶数题总分（Y）。将其带入公式：

$$R = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2]} \sqrt{[N \sum Y^2 - (\sum Y)^2]}}$$

最终 $R=0.91$ 。将 $R=0.91$ 代入 Spearman-Brown 公式进行校正：

$$R1 = \frac{2 \times R}{1 + R}$$

则 $R1=0.95$ ，整个测验的信度系数是 0.95，所以测试结果是

可靠的。

（四）效度

效度在评估语言测试中也起着至关重要的作用，不管是测试汇报还是在反馈时必然要考量的一方面。测试的效度指测试的有效度，即试卷题目考察的知识点是否考查了教学大纲中要求考查的内容。一般情况下，检验效度的通常做法是分析其测量结果和与之对应的有效标准，最后得出的相关系数越大表明试卷的效度就越好。根据王汉澜的《教育测量学》中，可以用下面的公式来表示效度与信度之间的关系：

$$R \leq \sqrt{R1}$$

该公式中，R 为效度系数，R1 为信度系数。

将 $R1=0.95$ 代入上述公式， $R \leq 0.97$ 。具有较强的相关性，比较理想。

四、试卷成绩分析

本研究从成绩基本描述和统计、成绩分布的频数及测试的成绩分布直方图等几个层面分析本次的试卷成绩。从这几个方面分析，可以比较直观地发现并看出试卷成绩的特点，也可以直接从直观图得知学生成绩的整体情况。

（一）成绩基本描述和统计

考试成绩的基本描述是分析班级成绩常用的一个方面，其指标通常包括：最小值、最大值、平均数、标准差、偏度系数等。通过 SPSS13.0 计算，本次试卷的输出描述统计如下图所示：

表 2. 本次试卷描述统计结果

	N	Minimum	Maximum	Mean	S.D	Skewness	Kurtosis
SUM	72	27	86.5	57.4	18.290	-.726	.279

由上图可知：本次测试中，班级的最低分 27 分，最高分 86.5 分，平均分 57.4 分，标准差 18.290。从这些数据中，可以看出该班学生成绩的大概情况。

（二）成绩分布的频数

测试成绩的频数分布的分析可以直观有效地体现出测试数据整体的分布情况。同时，频数分布的分析可以直观地看出考生的成绩主要集中在什么样的分数范围，它主要是通过平均分、中位数、众数、方差、标准差等指标来分析，代表的是考生的集体水平。

对本班成绩进行以上几个方面的检测，结果如下表所示：

表 3. 本班成绩多方面检测结果

本班平均分	57.4
本班中位数	58.5
本班众数	60
本班分数的方差	330.01
本班分数的标准差	18.12

通常情况下,我们通过描述这些分数的集中趋势,分析考试后数据的特征。在这些指标中,平均数最能体现班级学生的学习水平。中位数是按照顺序排列的数字中的中间的一个数,也是集中量的一个指标,也具备一定的参考价值。邹申指出:众数则是一组数值中出现频数最多的那个数,容易计算但是却不具备集中量的基本要求:准确和稳定(邹申,2011)。众数、中位数和平均数具有稳定性和不稳定的。在进行集中趋势分析时,不能将它们单独分开看待,要集中观察这三位数的数值。统计结果显示,它们相对比较接近,表明数据分布呈正态分布。

标准差也被称为标准偏差,是用来表示考试成绩最好的指标,能够更详细地描述一组分数的差异特征。标准差的数值和分数的离散程度呈现正相关。该组分数标准差是18.12,标准差相对来说较小,所以离散程度小。

(三) 成绩分布直方图

在本次检测中,成绩频数分布直方图如下:

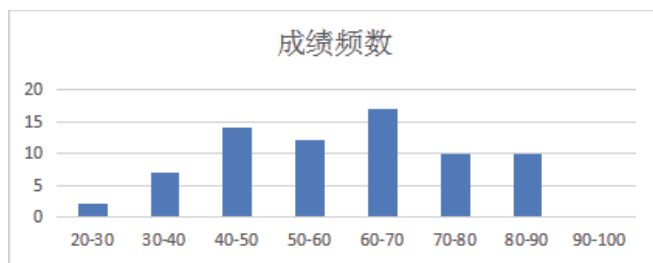


图1. 成绩频数分布图

由成绩频数直方图可知,60—70分区间的人数最多,有17人。其次人数多是40—50分区间,共14人;第三是50—60分区间的人数,共12人。

正常情况下,一个班的考试成绩应该是符合正态分布的,即低分段和高分段的同学比较少,中等成绩的同学较多。如果考试的结果不理想,低分数段的人数比较多,或者高分段的人数过多,这样的话,考试的最终结果就不会呈现正态分布。如果考试的结果没有呈现正态分布,则说明这套试题不能较好地反映出学生成绩的真实水平。根据SPSS13.0计算可得,如下图可知:本次测验中的本次考试分数基本呈现正态分布。

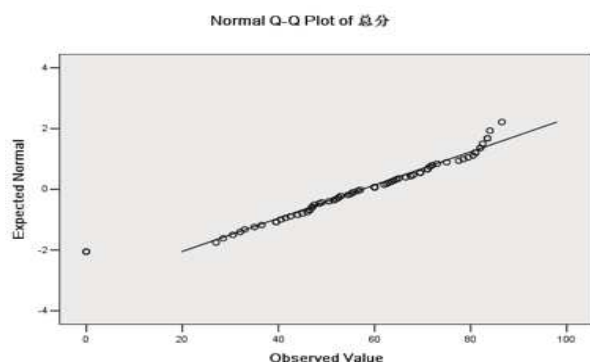


图2. 本次考试成绩正态分布图

五、教学反思和总结

本研究对本次测试试卷进行了多个维度的分析,得出的结论如下:该套试题的信度和效度均符合相关要求,相对可靠,所以该测试卷是一套比较合理的考试试卷。有些题目的难度适中,真实地反映出学生的真实水平。在本次检测中,共有72人参与考

试,其中及格人数37人,不及格人数为35人。其合格率大约为51.4%。学生的成绩基本呈现正态分布。本次测试中,最高分为86.5,最低分为27。平均分为57.4。但学生的整体成绩不太理想,学生还需要继续努力学习。由此,高分段和低分数段的水平还有很大的提升空间。

通过对试卷的分析,发现学生听力部分、阅读部分和翻译部分的得分率比较低。在今后的英语教学中需要特别注意听力部分、阅读部分和翻译部分的教学。在日常的教学过程中,要尽可能多从学生的弱势出发,创新教学方法,拓宽教学思维,挖掘好的英语学习资源,提高学生的英语综合应用能力。针对以上三点弱势模块,在以后的课堂教学中,要着重着眼于三点:首先,要更加重视听力板块的教学。要想提升学生的英语听力水平,课堂上要注重听力做题技巧的讲解和应用、理解和渗透及听力练习的布置,加强作业的反馈和练习的效率,要多给学生锻炼自己的机会,还要帮助学生找到一些适合他们并且有趣的听力材料、和学生日常生活相关的英文视频,在兴趣驱动和任务驱动的共同作用下,督促学生沉下心,多练习,以结果为导向;其次,要引导学生运用所学阅读技巧,尽可能多阅读英文材料。掌握精读和泛读的阅读技巧,养成阅读英文新闻和文章的习惯,长期监督和鼓励;最后,要多带领学生把翻译技巧运用到翻译文本中,选择合适并感兴趣的文本增加练习和纠错。学生只有经过大量的语言刺激和语言输入,才能会有一定量的输出,语言学习就要日积月累,量变达到一定程度,才能达到质的飞跃。

综上所述,通过本次大学英语测试的研究和分析,对课堂教学有了新的认识,积累了教学的经验,这为以后的大学英语教学工作指明了方向。

参考文献:

- [1] 陈露露. 国家英语能力等级考试研究[D]. 华中师范大学, 2019.
- [2] 高向琴, 董霞. 语言测试的反拨效应对西部高校大学英语考试问题及改革的启示[J]. 教育现代化, 2019, 6(66): 178-179+185.
- [3] 龚韶华. 语言测试的反拨作用与大学英语考试改革的思考——以呼伦贝尔学院为例[J]. 呼伦贝尔学院学报, 2021, 29(06): 34-37.
- [4] 贾国栋. 《大学英语教学指南》与高校大学英语教学改革[J]. 当代外语研究, 2017(06): 62-65.
- [5] 贾国栋. 大学英语口语测试的预期反拨效应——以全国大学英语四、六级口语测试为例[J]. 外语测试与教学, 2016(04): 1-9+51.
- [6] 赵辑. 基于文化资本理论的外语测试政策研究[D]. 上海外国语大学, 2017.
- [7] 金艳, 王伟, 杨浩然. 语言测试中的技术应用: 基于大学英语四、六级考试的实践分析[J]. 外语测试与教学, 2021(01): 1-7+27.
- [8] 金艳, 杨惠中. 走中国特色的语言测试道路: 大学英语四、六级考试三十年的启示[J]. 外语界, 2018(02): 29-39.
- [9] 苏悦畅. 浅析大学英语四级笔试的信度和效度[J]. 海外英语, 2019(16): 130-131.
- [10] 吕娜. 大学英语测试目前存在的问题及对策[J]. 外语学刊, 2016(01): 116-118.