

# 大学英语四级中阅读理解测试的效度研究

钟绍嫣

(北京大学 吉林省吉林市 132013)

**【摘要】**CET4是测量非英语专业大学生英语学习能力的有效手段之一。经过改革后的大学英语四级测试面有所拓宽, 题型有效性也有所提高。本文旨在通过对2019年CET4中阅读理解进行解析, 简要分析其信度和效度。

**【关键词】**大学英语四级; 效度研究; 阅读理解;

**DOI:** 10.12361/2705-0416-04-03-76955

自从英语成为中国常规测试中的三大主科以后, 中国英语考试便存在着应试教育的问题。在应试教育模式的培养下, 大多数学生思维模式固定, 忽视了英语思维的培养。升入大学之后, 非英语专业学习者不会将大量时间精力放在英语上。但随着英语教学模式的更新, 这一部分问题在近些年有所改善。

大学英语四级是能够测试中国大学生现有英语学习能力水平的考试, 其衡量英语测试的关键标准之二是信度和效度, 一份优秀的测试也离不开信度和效度的和谐与统一。本文旨在通过对 2019 年 CET4 试卷信度和效度的分析, 归纳有关于阅读题型的出题经验。

## 1 信度与效度

### 1.1 概念阐述

语言测试的信度与效度在国外一直有学者进行研究, 而在国内应用语言学领域, 从 20 世纪 90 年代测试的信度和效度的应用与研究才得到重视与发展。

信度、效度和操作可行性是衡量与检验语言教学质量的重要标准, 其中信度和效度是每一份试卷都需要接受反复考验与打磨的部分(杨惠中, 1998: 55)。信度是效度的前提, 一份试卷没有较高的信度必然不会带来相同等级程度的效度。

信度(reliability)指的是测试结果的可依赖性, 是一份试卷是否能够做到测试的一致性要求, 能够在多场测试中排除其他条件因素, 从而使得测试结果更加真实可靠、有依据性, 使得测试更加稳定。信度越高就能够说明考生受其他因素以及偶然性因素的影响越小, 更具有客观性。

效度(validity)是指测试的有效性, 一份试卷的测试内容和范围是否是应该被考到的, 是否在考生所掌握的知识范围之内, 是否能达到试卷所定的目标和目的。效度重点关注的是试卷中试题的编写是否是合理有效的(Messick, 1989, 转引自李清华, 2006)。效度是一项测试的核心部分。假如一项测试信度很高, 但试卷中应该进行考核的相应重难点都无从体现, 那这项测试效度就不高, 那这种类型的测试结果也就没有意义(董晓燕, 2008)。由此看来, 效度会比信度更好实施与测量。

### 1.2 信度与效度的关系

信度与效度与数学学科中常说的充分条件和必要条件相似, 两者可以同时存在但不是必须同时存在, 甚至可以说两者只要存在其一, 那测试在偏向于它的层面上就是成功的。一份优秀的试卷有时候是不能既满足信度又满足效度的。

#### 1.2.1 信度与效度相互对立

信度与效度相互对立。信度意味着测试的公平、客观和可实施性, 而效度则依赖于每场测试中考生的基本学习能力水平。某种意义上, 测试有了信度, 该测试的效度就不会太高。原因是在测试中照顾信度, 就无法兼顾效度。效度太高必将带来的是信度不足或较低。

#### 1.2.2 信度与效度相互依存

信度与效度相互依存。离开信度谈效度则丧失公正客观, 离开效度谈信度则缺少语言运用的实际环境和部分主观因素, 不利于研究的进行与测试的发展。所有任何一个测试及其题型都很难兼顾较高的信度和效度。研究者们努力通过大量的数据和样本进行分析, 研究效度和信度的高度结合。他们将整份试卷上的全部试题努力调整为较高级别的信度与效度的统一。笔者发现, 大学英语四级水平测试题型中, 阅读理解题型则满足了该理论研究方向。

## 2 大学英语四级阅读理解题型的相关分析

金艳(1998)曾以 CET4 为研究对象, 通过实验的方法研究阅读理解测试的效度。他以“自我内部反省法”(intro-spection)为研究手段, 将考生考试过程中的阅读行为进行一致性的分析, 以此来检验 CET 是否有效对考生的阅读能力进行考核。结果证明两者基本吻合, 因此说明了 CET 中阅读理解部分基本有效考核了大纲所规定的阅读技能, 如略读、寻读、仔细阅读等。

阅读材料可选择面广, 分数占比较大, 因此阅读理解是 CET4 中相比较为平衡的部分。也就是说只要能找到该语言测试中两者微妙的那个平衡点, 那么不仅能最大限度的保证该测试的客观性, CET4 测试的有效性也能随之发挥出最大效果和最大作用。

根据 2016 版《全国大学英语四六级考试大纲》来看, 新修订的大纲将考试内容作了进一步的调整。从四级笔试部分的测试题型和分值分布类型来看, 阅读理解约占整张试卷的 40%。阅读理解的三个部分又分别为: SectionA 选词填空、SectionB 段落匹配和 SectionC 的两篇阅读, 共三十道题。

由于客观题的答案是固定且唯一的, 同时采取了机器阅卷的方式, 从评分方式和评分结果看它保证了客观题的客观性, 具有较高的信度。但测试拥有较高信度的时候, 它的效度就无法被有效保证, 占比 30%的主观题部分由于答案不固定, 考生实际水平也不同, 该部分的信度难以保证。学生的回答反映出其掌握的英语知识多少和英语学习的实际水平, 该部分效度较高。

阅读理解共有四篇文章, 其中涉及的体裁类型多种多样, 考虑到考生的英语水平因素, 第一大类为科普类, 科普类文章占比几乎为一半, 其中可能会涉及到考生从未看过的单词, 但同时该测试也会贴心地在相应超纲单词后面缀上汉语意思以帮助考生了解文章。科普类文章涉及面较广, 且并不拘泥于科技领域, 在一般领域以及能够传达社会意义的领域也多有涉足。第二大类是社会生活方面的文章, 占比约 25%, 这一类文章大多贴近人们生活, 包括能够给人们生活带来便利的以及他们是如何发展的。考生在阅读过程中能够切身体会文章内容, 平时阅读经验较为丰富的考生甚至可以直接推导出文章大致意思。第三类就是教育、经济和文化三大层次方面的文章。

阅读理解不仅考察了句子层面的理解也考察到了语篇整体层

面的理解。就语篇理解而言,考生需要从上下文的语境中推测出大体意思,而句子理解方面则更偏向于仔细阅读技能的锻炼,考生需要通过问题所对应的句子及上下文进行详细分析,可能会涉及到同义替换或反义词的高级词汇改写。该部分考察了考生对句子的细节理解和对易混词语的掌握程度。相对于主观题和其他题型来说,阅读理解的信度和效度在一定程度上可以兼顾,客观题的机器评分更加方便省时省力,能够被广泛使用。因此阅读理解测试的信度和效度是有所保障的。

由此看来,考生们应该在日常生活中加大阅读量,拓宽阅读面。根据考纲规定,大学英语四级的词汇量应该掌握在 6000~8000 个左右,规定的每分钟阅读词汇应为每分钟 70 个词,而四级考试中规定时间为 40 分钟,实际上四篇文章以及问题的词量并不少,考生要在 40 分钟之内快速读完并做出选择。没有较深的功底及单词量支撑,的确有些困难。没有较快的阅读速度和较大的单词量,考生势必会在阅读中心烦气躁,也有可能将较简单的内容选错。

选词填空其实是之前旧题型完型填空的变体。在我看来,难度是有所提升的,完型填空所给选项时,可以看出该空缺需要填写的是什么词性的单词,但是选词填空则更需要考生自己根据前后文或文章整体意思来填写相关单词。选词填空要求考生有更加敏锐的判断和把握文章整体的能力。

段落匹配其实是将阅读理解中主旨大意题部分抽取出来,单独成题,主要锻炼的是考生略读技巧以及概括文章主旨大意的能力。SectionC 部分的两篇阅读主要考察的是考生的综合能力,每篇文章对应五个选择题,既有细节题,锻炼考生仔细阅读的能力,也有相关延伸表明类题目,能为考生带来较多的思考和启发。

以 2019 年 12 月大学英语四级考试题第二套为例:

#### 1.Banked Cloze

选词填空部分选取的材料是有关于机场和旅游的,进而谈到了感染风险。比较贴近人们生活,也能够让考生快速进入阅读理解状态,在面对第一篇阅读理解材料的时候不至于因为材料太难而上来就产生退缩心理。长度在 200 字左右,篇幅不长,符合考纲的规定。从题型上看,都是从 15 个备选词汇中选出 10 个最为匹配的词汇。参照《教学要求》附录、大学英语参考词汇表来判断词汇的难易程度,卷二考察了 15 个一般词汇。考生们能够掌握简单词汇,这部分并不会对考生产生差异性。

#### 2.Passage Reading

《考试大纲》规定 SectionC 部分阅读包括阅读短文两篇,每篇长度为 300~350 个词。从谈论话题上看,涉及教育资源,工厂工

作及资源问题,第二篇文章说的是 sea slug,部分学生可能没有听过,但是并不耽误考生们在文章中寻找答案。基本上都是学生较熟悉的体裁和话题,符合《考试大纲》的要求。

从内容效度上分析,该题型并没有直接考察考生的实际语言应用能力和表达能力,这一点和《教学大纲》中所规定的教学目标有差距,与快速阅读考察重点又有所重复,一定程度上减弱了四级考试的整体效度。

#### 3.Skimming, Scanning

段落匹配给出的标题是“Is Breakfast the Most Important Meal of the Day?”,主旨大意给出的很清晰明了,早餐是一天中最重要的一餐吗?其中包括吃早餐的好处,不吃早餐的坏处,以及研究者、研究机构等较为容易找到的部分供考生们寻找。段落匹配篇幅较长,涉及的段落也很多,其中又有较多信息是交叉重叠的,容易造成混淆,同时 1000 字左右的篇幅很容易让考生不知所云。但整体看来该篇段落匹配不难,信息较清晰。

从内容效度分析,快速阅读是从文中快速找出 10 个信息的段落来源,主要考察的是考生如何运用略读和查读技能查询信息的能力,这与《考试大纲》的规定一致,但与《教学要求》的教学目标有差距,影响了考试效度。由于考试对教学的反拨效应,四级考试中测试题型内容单一、与教学目标存在差距的现实,会削弱大学英语有效教学的实现。

通过分析大学英语四级考试 2019 年 12 月卷二部分的阅读理解文章和题目,发现选词填空信度较高,短文阅读信度差异较明显,快速阅读信度较为稳定。从内容效度看,选词填空、短文阅读对英语运用考察较单一。清一色地采用选择题的形式会影响对考生英语综合运用能力的考察,效度受到影响。

### 3 结语

总体来说,对比《考试大纲》来看,四级阅读理解信度较好,按照《教学要求》,四级阅读理解的效度还应该有所提高。如果能对阅读理解的题目类型做进一步的调整,设计一些主观题进行阅读测试,该部分测试就能够更加客观、有效地测量出非英语专业大学生的语言综合运用能力,从而提高 CET4 的信度与效度,使其更好的服务于考生服务于测试,从而进一步开展大学英语的有效教学。

作者简介:钟绍嫣(1998.12—),女,吉林省吉林市人,研究方向:英语教育。

### 【参考文献】

- [1] 陈巧巧.试论大学英语测试的信度和效度[J].青年文学家,2013(36):178-178.
- [2] 董晓燕.大学英语语言测试中信效度浅析[J].科技资讯,2008(34):211-211.
- [3] 刘润清,韩宝成.语言测试和它的方法[M].北京:外语教学与研究出版社,1991.
- [4] 吴小珊.论英语测试信度和效度的平衡关系[J].广西民族大学学报(哲学社会版),2007,29(2):129-130.
- [5] 全国大学英语四、六级考试委员会.大学英语四级考试大纲(修订版)[M].上海:上海外教教育出版社,2006.
- [6] 杨慧中.关于语言测试学科发展的若干思考[C].庄智象.中国外语教育发展战略论坛.上海:上海外语教育出版社,2009.
- [7] 金艳,吴江.以“内省”法检验 CET 阅读理解测试的效度[J].外语界,1998(2):47-52.
- [8] 教育部高等教育司.大学英语课程教学要求[M].北京:高等教育出版社,2007.
- [9] Oller J. Language Tests at School[M]. London: Longman, 1979.
- [10] Messick S. Validity and washback in language testing[J]. Language Testing, 1996(13):241-256.
- [11] Bachman L. & A. Palmer. Language Testing in Practice[M]. Oxford: OUP, 1996.
- [12] Bachman, L.F. & A.F. Secondary Language Testing in Practice[M]. Oxford: Oxford University Press, 1996.