

## 词汇测试应考虑的几个关键问题

王源

泰山护理职业学院 山东泰安

**【摘要】**词汇在语言习得中的重要性不言而喻，词汇测试也被广泛用来衡量学习者的整体语言水平以及探讨其与听说读写等分项语言技能之间的相关程度；词汇习得在语言习得的整个周期中占有很大比例，需要学习者投入大量的时间和精力，因此词汇习得的研究仍需要学界的广泛关注。词汇具有多维的特点，测试工具和方法也日趋多元化；对词汇测试结果的解读应考虑一些具体细微的指标。本文基于国内外近二十年来对词汇测试的研究，梳理了以往研究的概况，指出了这些研究中存在的不足。本文从词汇量的具体划分、历时与共时测试、测试词的选取、测试方法和计量单位等五个方面进行了较为详细的论述，旨在更好地发挥词汇测试对于词汇教学的反拨作用，提高词汇教学的效率。

**【关键词】**词汇测试；语言水平；学习者；语言技能

**【基金项目】**泰安市哲学社会科学规划项目，课题编号 18skx122

**【收稿日期】**2024年11月13日

**【出刊日期】**2024年12月26日

**【DOI】**10.12208/j.ije.20240114

### Some key issues in vocabulary test

Yuan Wang

Taishan Nursing Vocational College, Tai'an, Shandong

**【Abstract】** It is widely acknowledged that vocabulary plays a vital part in language acquisition. Vocabulary test has been widely adopted to evaluate learners' overall language competence and correlation to specific language skills such as listening, speaking, reading and writing. Vocabulary acquisition takes up a big proportion in the whole cycle of language acquisition and learners need to spend much time and energy on it, therefore studies in this field still deserve much attention. As vocabulary is a multi-dimensional construct, test tools and methods are becoming increasingly diversified. The outcome of vocabulary test should be interpreted taking some minor indexes into consideration. Based on studies in the past 20 years, this paper first briefly summarizes their major findings and points out the inadequacies. Then this paper gives a detailed statement of vocabulary test including five aspects: specific indexes of vocabulary size, diachronic and synchronic tests, selection of test items, test methods and units of counting. This paper aims at improving the backwash effect of vocabulary test on vocabulary teaching and raising the efficiency of vocabulary teaching.

**【Keywords】** vocabulary test; language competence; learner; language skills

#### 1 引言

词汇量与各项语言技能有着不同程度的相关性，与综合语言能力也有一定的关联，长期以来成为一个研究热点，国内研究机构和学者对不同群体的词汇量也进行过多次测试，以期期为教学提供诊断和参考；国内对词汇量的测试始于二十世纪八十年代，兴盛于二十世纪九十年代和二十一世纪初，主要测试学习者的被动词汇量，探讨其与分项语言技能或者综合语言能

力的关系；进入二十一世纪，产出性词汇量的测试引起了一些学者的关注，这一部分研究主要探讨主动词汇与写作质量的关系（刘东虹 2004）<sup>[1]</sup>；少部分研究关注口语词汇量（甄凤超 2005）<sup>[2]</sup>。

#### 2 词汇测试应关注的几个关键点

词汇测试的意义在于为语言教学提供参考或者是揭示其与综合语言能力的关系，然而，综合语言能力的测试比较困难，不同的具体语言技能对词汇的具体要

求也有很大不同,如阅读技能主要对视觉词汇量有较高要求,而听力则主要依赖于听觉词汇量,写作则要求有较高的自由主动词汇量。然而,测试结果却存在着较大的差异,这些差异的原因主要是在计数单位,词类,词频分布以及测试方式的不一致造成的。本文拟通过对以往词汇量测试时没考虑周全的一些细节进行分析,以期对今后的词汇量测试提供一定的参考。词汇测试要考虑词汇量的具体划分,历时和共时,测试词的选取,测试工具的使用,计量单位以及词汇量计算公式等。词汇量与词汇深度二者相互关联,词汇量大的学习者通常能掌握和使用更多的低频词,而这被认为是词汇深度的一个重要标志。

### 2.1 词汇量的具体划分

国内的词汇量测试主要是阅读词汇量,也称为消极词汇量,接受性词汇量,被动物词汇量或认知词汇量。

(邓昭春, 2001)<sup>[3]</sup> (吕长竑 2004<sup>[4]</sup>); 指的是学习者能通过视觉识别通达其词义。而反映学习者口语和写作中词汇使用情况的积极词汇量,或称为主动性词汇量、产出性词汇量的研究相比却较为匮乏。需要注意的是,主动性词汇还可以进一步细分为控制性主动词汇(controlled active vocabulary)和自由主动词汇(free active vocabulary)(Laufer 1998)<sup>[5]</sup>,其中控制性主动词汇指的是任务激励下可以产出的词汇,而自由主动词汇指的是学习者在没有任何激励的情况下能自由产出的词汇。而直接测量学习者的自由主动词汇是极其困难的,大多数的测试是通过限时作文来获得数据。崔艳嫣 王同顺(2006)<sup>[6]</sup>采用限时命题作文对学习者的产出性词汇进行了研究,但是也存在一定的局限性:命题作文的词汇往往围绕某一主题,不能反映学习者产出性词汇的概貌。

以上两类词汇量主要是反映学习者阅读和写作中的词汇量指标,而听力和口语中的词汇量研究相对比较匮乏,由于中国英语学习者听说的机会远远少于本族语者,通常认为他们的听力词汇量和口语词汇量的发展落后于阅读词汇量和写作词汇量。甄凤超(2005)<sup>[2]</sup>采用语料库和对比分析的研究方法,发现中国学习者的口语词汇量较少。

### 2.2 词汇量测试的历时与共时

词汇习得具有动态性和递增性的特点,同时在习得过程中也不可避免地伴有磨蚀现象,因此词汇量不是一个静态的数据,而是处于不断发展变化中。对词汇量的测试固定在一定时间点固然有其意义,但是不能描述词汇发展过程,历时的研究能更好地追踪学习者

的词汇获得(acquisition),磨蚀(attrition)和保持(retention),为外语教学提供更丰富,更深入的参考依据。国内近二十年的词汇量研究大多是静态的测量,历时的追踪研究较少,周大军,文渤燕(2000)<sup>[7]</sup>对理工科学生在校四年的词汇量进行了跟踪研究,发现词汇量的发展不是线性的,而是表现出先增后减的趋势;谭晓晨(2006)<sup>[8]</sup>选取了英语专业四个年级学生考察其产出性词汇发展状况,收集了一段时间内学习者的作文语料,发现产出性词汇发展不是单一,线性的,而是呈现“先快后慢”的特点,存在词汇高原现象;谭将产出性词汇量发展分为迅猛阶段,过渡阶段和维持阶段。崔艳嫣,王同顺(2006)<sup>[6]</sup>同样选取了英语专业四个年级学生为研究对象,对其接受性词汇量和产出性词汇量的发展路径进行了研究,发现接受性词汇量呈线性增长模式,产出性词汇量的发展存在“高原现象”。需要指出的是,这一研究选取的不是同一群体,而是对四个不同年级的大学生进行横向比较,不属于真正意义上的追踪研究,所揭示的词汇量发展特点值得斟酌。卢敏(2008)<sup>[9]</sup>对中国学生英语口语语料库中笔语部分的作文语料中英语专业一至四年级学生的议论文作文进行了分析,也得出了类似的结论,即产出性词汇的发展存在高原现象,二年级学生的产出性词汇达到最高阈值,三、四年级学生则停滞不前。

### 2.3 测试词的选取

由于英语中的词汇浩如烟海,把每个词都拿来对学习者的测试显然是不现实的;通常的做法是选取一定数量的词来进行测试,测试词的选取对于词汇量的计算结果有着直接影响。选取时通常要考虑取词类型,抽样方法,词频分布,词类,测试词的数量,词表等因素。

#### (1) 取词类型

多词单位(multi-word unit)也是词汇的重要组成部分,应该纳入词汇量测试的范围;然而实际上却是绝大多数研究都忽略的一方面。

#### (2) 抽样方法

抽样方法包括词频法和词典法两种,其中以第一种方法居多(桂诗春 1985<sup>[10]</sup>),主要因为后者存在一些不足:词典中的词条占据的篇幅不等,多义词占据的篇幅相对较大,因此从这个角度而言词典的等距离抽样实际上是不等距离抽样;同时,多义词大多为高频词,被试认识这些词的可能性较大,应避免对这些词的过度抽样,否则易造成对被试词汇量的过高估算。可采用给单词编号,每隔固定数量的单词进行抽取,或者采用

随机取样、分层抽样的方法进行抽取；桂诗春(1985)<sup>[10]</sup>采用分层随机抽样，周大军，文渤燕(2000)<sup>[7]</sup>采用随机等距抽样。

### (3) 词频分布

不同水平学习者掌握的词汇具有不同的词频分布，所选测试词的词频分布应尽可能覆盖学习者所掌握的词频层级。

### (4) 词类

词汇测试所选词类主要为实词类，其中以名词，动词，形容词居多。测试词中这三类词应具有合理的比例，然而一些研究未考虑这一问题，所选词类过于单一或者词类之间的比例不合理；通常按照名词:动词:形容词=5:3:2。

### (5) 测试词数量

词汇量的测试都是以有限的词汇估算学习者的词汇量总数，是一种粗略的估计（桂诗春 1985）<sup>[10]</sup>，测试词越多，测试结果的可信度越高，在一定的测试时间内，测试格式（test item format）决定了可测试词的数量。

研究者, 年份	测试词数量
桂诗春 1985 <sup>[10]</sup>	200
周大军, 文渤燕 2000 <sup>[7]</sup>	200
吕长竑 2004 <sup>[4]</sup>	100
崔艳嫣, 王同顺 2006 <sup>[6]</sup>	156

### (6) 词表

英语通用词表 *A general Service List of English Words*(GSL),收录了 2,000 个核心词(core vocabulary),《大学英语教学大纲通用词汇表》收录了 6 个等级的《大学英语教学大纲(文理科本科用)》的词汇表。《大学英语教学大纲》(1984)包括 3820 个单词,分为 0-4 级。

## 2.3 测试方法

何谓“认识一个单词”，测试的方法也存在很多差异。按照测试手段可以分为纸笔测试和计算机测试两大类。

### (1) 纸笔测试

国内的词汇量研究大多采用纸笔测试（桂诗春 1985<sup>[10]</sup>，周大军，文渤燕 2000<sup>[7]</sup>，吕长竑 2004<sup>[4]</sup>）

#### 1) 客观题型

##### 四选一

很多国内学者采用四选一的题型，要求被试为英语单词选择正确的中文释义；这实际上考查的是形式

意义匹配(form-meaning match)。桂诗春(1985)<sup>[10]</sup>，周大军，文渤燕(2000)<sup>[7]</sup>，吕长竑(2004)<sup>[4]</sup>均采用此方法。

### Vocabulary Levels Test(VLT)

国外使用最广泛的词汇量测试是词汇水平测试，近年来国内学者也开始使用；该测试要求被试将英文释义与英语单词匹配，每组里包括 3 个英文释义和 6 个英语单词备选项，示例如下。李晓(2007)<sup>[11]</sup>在此基础上设计了一套试题对非英语专业大学生的词汇量进行了测试。

### Vocabulary Size Test(VST)

该测试题干为一目标词和一个非定义句子，选项为四个英文释义，要求被试为目标词选择正确的英文释义。

### Checklist Test

这一测试方法让被试读词表，判断是否认识这些词条；因此又被称为 Yes/No 测试，这实际上是一种意义回忆(meaning recall)测试。该测试手段简单易行，但是存在受主观因素影响较大等缺点。

### 2) 主观题型

为英语单词写出中文释义

### The Productive Vocabulary Levels Test(PVLT)

该测试形式为句子填空，为避免同义词等其它可能的答案，目标词的第一个或前几个字母给出。

Paul Meara(2005)<sup>[12]</sup>采用蒙特卡罗模拟方法认为 LFP(Lexical Frequency Profile)在比较具有词汇量本质性差异的水平组之间能发挥最佳作用，而在考察词汇量的适度变化时则可能“力不从心”。国内研究者多采用这一版本，即把词汇分为四类（卢敏 2008）<sup>[9]</sup>。

### 3) 量表

量表测试可看作客观和主观题型的结合，被试既需要做选择题，又需要做一些填空甚至造句。Paribakht&Wesche(1997)<sup>[13]</sup>开发了词汇知识量表(Vocabulary Knowledge Scale-VKS)，该量表分五个等级，涵盖“没见过这个词”，“见过这个词但不知其义”，“见过这个词且知其义”，“确切认识这个词且知其义”和“能正确使用这个词造句”。，崔艳嫣，王同顺(2006)<sup>[6]</sup>，李晓(2007)<sup>[11]</sup>先后直接使用该量表或者基于此量表做一些修改进行词汇知识的测试。

### (2) 计算机测试

随着计算机技术的发展，交互式的计算机化测试日益成为可能。Laufer&Goldstein(2004)<sup>[14]</sup>设计了 *Computer Adaptive Test of Size and Strength(CATSS)*，该测试工具具有两个较为明显的优势：1. 根据被试的测

试情况可进入到更高一级的词频水平,如被试在 1,000 词频水平表现优异,则程序将测试进入到 2,000 词频水平,以此类推直至被试出现较多的错误。2. 相比静态测试,这一测试工具能够避免在太难或太易的单词上浪费时间,达到更好的抽样效果;还可以避免传统测试中实施者对被试词频分布情况所做的主观猜测,提高测试的效度。

#### 2.4 计量单位

计量单位的确定对于词汇量测试结果有着直接的影响,不同的计量单位会导致迥然不同的结果。目前可供选择的计量单位包括类符,词目和词族等。计量单位的确定对于词汇量测试结果有着直接的影响,不同的计量单位会导致迥然不同的结果。近年来,国外学者提出基础词(base words)和词族(word families)两个概念。基础词指的是词典收录的词目形式,如 care, careful, careless 在一般词典中为独立词条,属于基础词;而 carefully, carelessly 则不属于独立词条,不应计算为基础词。

计量单位的选取需要基于研究目的, Norbert Schmitt(2015)<sup>[15]</sup>认为:对于一语儿童,按照类符计算较为适宜;对于成年本族语人,词目是比较好的计量单位。对于非本族语人,情况则较为复杂。他们的语法词缀能力可能差距很大,对于初学者来说,未必能掌握同一词目中的所有成员,因此类符可能是较好的计量单位。对于程度较高的学习者来说,按照词目计算则更佳。词族作为计量单位也有其自身优势:与词典中的基础词对应密切,可以避免过度计量。Paul Nation(2001)<sup>[16]</sup>认为:计量单位的选取还要考虑接受性产出性方面。接受性方面词族是最佳计量单位,产出性方面词目甚至是类符可作为最佳计量单位。Norbert Schmitt(2015)<sup>[15]</sup>认为:总体而言,词目可能是最佳计量单位;计量单位的选取要反映研究目标,被试语言水平和研究来源等。

#### 3 结语

词汇量曾经被学习者神化,认为是衡量一名学习者英语水平的最重要指标之一。其实不然,因为“量”和“质”要均衡发展,片面追求量的增长会导致对高频词的掌握“一知半解”,“不求甚解”,而低频词“学而不用”,“鲜有用武之地”。词汇的使用频率有很大差异,不能平均用力,对“性价比”高的词汇要舍得投入“时间代价”来精学多练,力求全面地掌握这些词汇的“性能”;而对于“性价比”低的词汇适可而止为宜。

词汇测试归根结底是为外语教学提供有价值的参考,因此词汇测试一定要客观真实地反映学习者的真

实面貌;这就对测试的信度和效度提出了很高的要求。在大数据,云计算等先进技术的推动下,我们有充分理由相信这一领域的研究将取得更多丰硕的成果。

#### 参考文献

- [1] 刘东虹. 写作策略与产出性词汇量对写作质量的影响[J]. 现代外语, 2004 (3).
- [2] 甄凤超. 中国学习者英语口语词汇量及常用词汇研究——基于英语口语语料库的词目研究[J]. 解放军外国语学院学报 2005 (5).
- [3] 邓昭春. 英语词汇量调查问题探讨——兼评一份全国词汇量调查表[J]. 外语教学与研究, 2001 (1).
- [4] 吕长竑. 词汇量与语言综合能力、词汇深度知识之关系[J]. 外语教学与研究, 2004 (3).
- [5] Laufer, B. The Development of Passive and Active Vocabulary in a Second Language: Same or Different? [J]. Applied Linguistics 19,3: 255-271.1998.
- [6] 崔艳嫣 王同顺. 接受性词汇量、产出性词汇量与词汇深度知识的发展路径及其相关性研究[J]. 现代外语, 2006 (4).
- [7] 周大军 文渤燕. 理工科学生英语词汇量状况全程调查[J]. 外语教学与研究, 2000 (5).
- [8] 谭晓晨. 英语词汇深度知识习得过程初探——一项基于词义与搭配的研究[J]. 解放军外国语学院学报, 2006 (3).
- [9] 卢敏. 产出性词汇知识广度的发展特征——基于英语专业学生书面语的研究[J]. 外语教学理论与实践, 2008 (2).
- [10] 桂诗春. 我国英语专业学生英语词汇量的调查和分析[J]. 现代外语, 1985 (1).
- [11] 李晓. 词汇量、词汇深度知识与语言综合能力关系研究[J]. 外语教学与研究, 2007 (9).
- [12] Meara, Paul. 2005. Lexical Frequency Profiles: A Monte Carlo Analysis. Applied Linguistics 26/1:32 - 47.
- [13] Paribakht, T. S.&Wesche, M. (1997). "Vocabulary enhancement activities and reading for meaning in second language vocabulary development" in second language vocabulary acquisition. Cambridge: Cambridge University Press.
- [14] Laufer, B., & Goldstein, Z. 2004. Testing vocabulary

knowledge: Size, strength, and computer adaptiveness.  
Language learning, 54(3), 399-436.

- [15] Norbert Schmitt. 2015. 词汇研究[M]. 北京: 外语教学与研究出版社
- [16] Nation, P. 2001. Learning vocabulary in another language. Cambridge: Cambridge University Press.

**版权声明:** ©2024 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。  
<http://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**