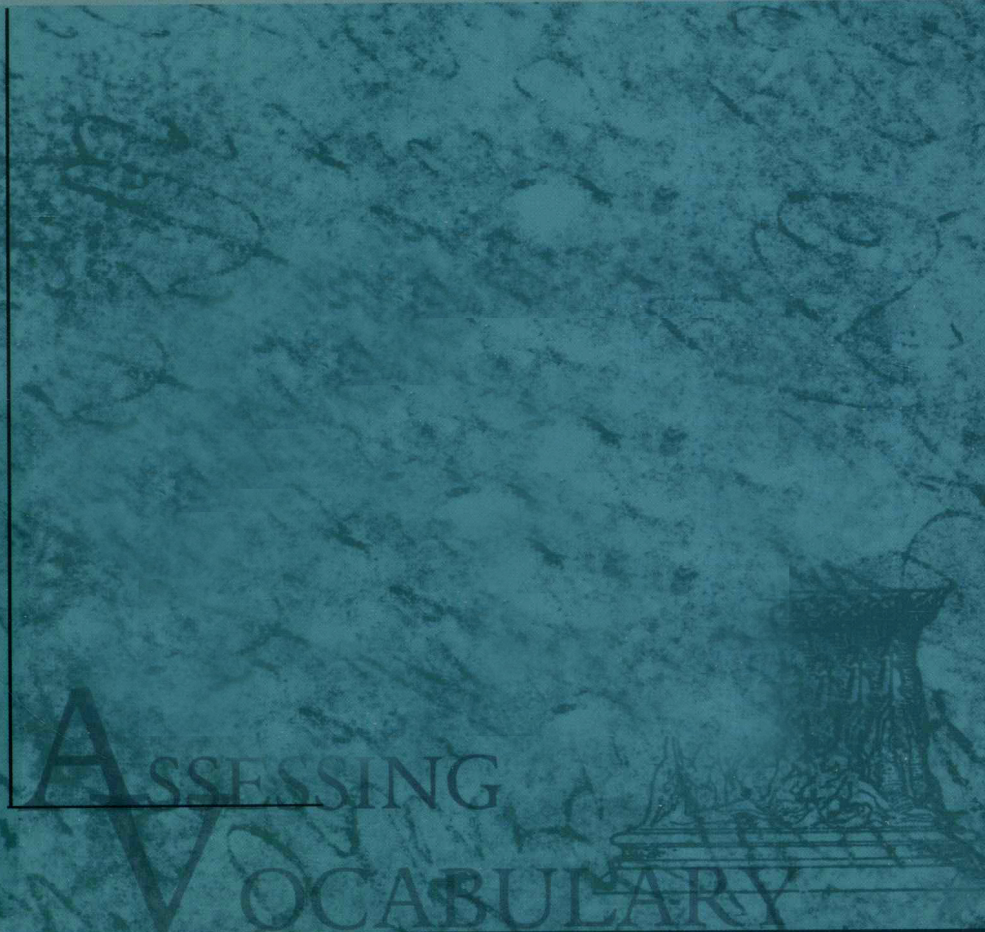


Assessing Vocabulary.....

词汇评价

John Read (新西兰) 著



CAMBRIDGE

E n g l i s h T e a c h e r

..... 剑 · 桥 · 英 · 语 · 教 · 师 · 丛 · 书

Assessing Vocabulary.....

词汇评价

John Read (新西兰) 著
何安平 导读

外语教学与研究出版社 北京

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS BEIJING

剑桥大学出版社

CAMBRIDGE UNIVERSITY PRESS

京权图字：01-2010-6798

This is a reprint edition of the following title published by Cambridge University Press:

Assessing Vocabulary (ISBN: 9780521627412)

© Cambridge University Press 2000

This reprint edition for the People's Republic of China (excluding Hong Kong SAR, Macao SAR and Taiwan Province) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Foreign Language Teaching and Research Press and Cambridge University Press 2010

This reprint edition is authorized for sale in the People's Republic of China (excluding Hong Kong SAR, Macao SAR and Taiwan Province) only. Unauthorized export of this reprint edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of both Cambridge University Press and Foreign Language Teaching and Research Press.

本书版权由剑桥大学出版社和外语教学与研究出版社共同所有。本书任何部分之文字及图片，如未获得两社书面同意，不得用任何方式抄袭、节录或翻印。
只限中华人民共和国境内销售，不包括香港特别行政区、澳门特别行政区及台湾省。不得出口。

图书在版编目(CIP)数据

词汇评价 = Assessing Vocabulary: 英文 / (新西兰) 里德 (Read, J.)
著; 何安平导读. — 北京: 外语教学与研究出版社, 2010.10
(剑桥英语教师丛书)
ISBN 978-7-5135-0245-0

I. ①词… II. ①里… ②何… III. ①词汇—评价—英文 IV. ①H03

中国版本图书馆 CIP 数据核字 (2010) 第 198062 号



悠游网—外语学习 一网打尽

www.2u4u.com.cn

阅读、视听、测试、交流、共享

封底刮刮卡，获积分！在线阅读、学习、交流、换礼包！

出版人：于春迟

项目负责：范海祥

责任编辑：范海祥

执行编辑：郑加丽

封面设计：袁璐

出版发行：外语教学与研究出版社

社址：北京市西三环北路19号(100089)

网址：<http://www.fltrp.com>

印刷：三河市北燕印装有限公司

开本：650×980 1/16

印张：19.5

版次：2010年12月第1版 2010年12月第1次印刷

书号：ISBN 978-7-5135-0245-0

定价：38.00元

购书咨询：(010)88819929 电子邮箱：club@fltrp.com

如有印刷、装订质量问题，请与出版社联系

联系电话：(010)61207896 电子邮箱：zhijian@fltrp.com

制售盗版必究 举报查实奖励

版权保护办公室举报电话：(010)88817519

物料号：202450001

出版说明

随着国家《英语课程标准》的实施以及《国家中长期教育改革和
发展规划纲要》(2010—2020 年)的颁布,英语教育领域正在进行着全面
而深入的改革。新课程改革要求教师提高素质、更新观念、转变角色,
同时要求教师的教学行为发生相应的变化。研究学生如何学、教师如何
教、考试如何考和教师如何在教学实践中提高自我素质已经成为目前英
语语言教学研究的焦点。因此,教师培训工作的重要性和关键作用变得
越来越突出。为了更好地配合新课程改革,提高教师的自我素质,外语
教学与研究出版社精选并引进了剑桥大学出版社设计的这套“剑桥英语
教师丛书”。

这套丛书由当今国际语言教育界的权威人士 Gary Buck, John Read,
J. Charles Alderson 等编著,由国内英语教育界专家、学者撰写导读。这
些专家、学者在认真研读原著的基础上,结合中国外语教师的需求和现
状,有针对性地用深入浅出的语言对原书作了概括、总结与评价,有助
于英语教师及英语专业的研究生对原著的学习与理解。

本丛书的选题都是目前英语教师关注的教学实践中的热点专题。有
几本是关于英语教学中的测试方面的,如 *Assessing Vocabulary*, *Assessing
Listening*, *Assessing Speaking*, *Assessing Reading*, *Assessing Writing*; 有
关于语言教师开展自我研究的,如 *Psychology for Language Teachers*; 有
关于将信息技术用于语言教学之中的,如 *Language Learning in Distance
Education*, *Assessing Language through Computer Technology*。

本丛书从英语语言教学的理论和实践两个层面为英语教师的教学实践提供理论指导，并为英语教师的在职教育和终生职业发展提供丰富的资源，可以作为英语教师专业化教育与发展和英语专业研究生教育的教材，供在职英语教师和英语语言教学研究者，尤其是英语专业的研究生学习使用。

外语教学与研究出版社

词

汇评价

导 读

《词汇评价》是一本关于二语学习中词汇评价的专著，作者是新西兰维多利亚大学应用语言学院的教授 John Read 博士。词汇研究已成为当今应用语言学和语言教学界的热门话题。作者在全面回顾二语教学中词汇的本质、词汇习得、词汇应用等研究的基础上，对词汇评价领域的传统研究在理论和实践层面上进行了全面介绍。他通过研究大量个案和分析典型案例，对当今流行的各种词汇测试模式进行了介绍和评论，包括各自的测试理念、形式特点、用途和有效性；还对未来的词汇测试研究，尤其是语料库语言学视角下的词汇研究与测试的最新发展趋势作了深入探讨。

1

全书结构和总体评价



全书结构

本书是剑桥语言评价系列专著中的一部，初版于 2000 年，到 2009 年已经是第 8 次重印。著名语言测试专家，也是本丛书的主编 J. Charles Alderson 和 Lyle F. Bachman 为本书联合作序，接着是作者本人的致谢和全书简介。正文部分共 8 章，251 页：（1）词汇在语言评价中的地位；（2）词汇的本质；（3）词汇习得和使用的研究综述；（4）词汇评价研究的综述；（5）词汇测试：四个经典案例分析；（6）分离式词汇测试设计；（7）综合型词汇测试；（8）词汇评价的发展趋势。正文之后还有参考文献和索引共 28 页。



总体评价

本书作者多年来从事二语教育环境下的词汇教学和词汇评价研究，位居该领域前沿。本书特点至少有三：一是展示了如何将二语词汇习得的研究成果应用于二语词汇评价；二是使用案例分析和评价了词汇评价的各种方式；三是将语料库语言学这门新兴学科在词汇研究方面的理念和

方法应用于二语词汇评价，并视其为该领域发展的重要趋势。作者在本书中构建了词汇测评的总体框架（包括两大趋势和三个维度，详见下文），然后以此统领全书的撰写。书中既有对传统词汇测试理念和方法的回顾与反思，也有对当今测试新理念与研究成果的介绍和评价。书中介绍的关于词汇的深度知识测试、词汇在真实语境中的意义理解与应用、测试内容和考生答题的语域特征等重要理念，对外语词汇教学、测试命题者和研究者均有启发和指导作用。因此，两位为本书作序的语言测试专家称赞本书的出版很及时，有权威性、前瞻性、建设性和实用性¹。

¹ 本文引用人名和出处均出自原著，故不再附带参考文献。

2

各章节主要内容与导读

●● 第一章 词汇在语言评价中的地位

本章主要介绍语言测试的两大趋势以及词汇评价的三个维度。作者首先指出，词汇在二语教学中的优先地位使之成为检测学习者语言知识和能力的主要手段。接着他介绍了语言测试的两大趋势：一是分离式途径；二是交际式途径。前者又称传统测试方法，它先将目标语的知识结构进行分离，然后逐一测试学习者对知识的了解，例如单词的意义、词形、句型、发音区别等。而后者则是近 30 年来以 Bachman 和 Palmer (1996) 为代表的学者倡导的交际型测评模式。它旨在通过任务型的测试题同时考查语言知识和语言策略使用两个方面，并由此推导学习者的整体语言运用能力，词汇知识仅仅是其中一个微小的组成部分；而且词汇知识也不仅仅指发音、拼写和孤立的意义，而是要扩展至与之相关的语法、语篇、功能和社会语言学知识（见 P. 6 的框架表）。

接下来作者阐述了词汇评价的三个维度：一是分离式还是嵌入式；二是选择型还是综合型；三是脱离语境还是

依赖语境。其中维度一关注的是词汇知识是独立或平行于语言系统的能力，还是某种语言能力的下属成分。例如：在评价学术写作能力或阅读理解能力时仅将词汇列为评分指标之一，而非单独出一套词汇测试题。维度二关注的是测试所涵盖的词汇的广度，其中选择型是指命题者从已有的单词表中抽选一部分作为考查目标，考查学生对这部分词的意义和用法的掌握情况；而综合型则关注考生在语言产出中词汇使用的整体情况，例如总词汇量、词汇的多样性、复杂词和低频词的使用情况等。维度三关注的是语境在词汇测试中的作用。传统的语境观是为测试的词汇提供一个句子作为语境；如今的语境观则超越了句子层面而扩展至整个语篇，例如：要求考生在更宽泛的上下文语境中判断某个词在该语篇中的意义而非其孤立的意义。

以上三个维度虽然不足以覆盖词汇测评的全貌，但却能为目前各种各样的评价模式提供一个基本的分类评估框架。至此，作者初步构建了语言测试的两大趋势和词汇评价的三个维度，并将在其后的各章中使用这个总体框架来论述和评价各类词汇习得研究、各种评价模式及具体案例。

●● 第二章 词汇的本质

本章讲词汇的定义、词汇知识构成和词汇能力构成三个大问题。

什么是词汇 (vocabulary)？以往的研究认为词汇由单个词 (individual word) 汇集而成，每个词都有独立的意义，尤其是字典意义。然而，当代语言学研究，尤其是语料库语言学的研究对于词这个语言的基本意义单位却有了一些新思考。例如区分词型 (type) 和词符 (token)，将词型和词符的比率 (type-token ratio)、词目 (lemma)、词族 (word family)、词项 (lexical item)、词汇短语 (lexical phrase) 和多词词项 (multi-word item) 等概念用于词汇研究。每一种观点的提出都深化了人们对语言意义表达单位的认识，而且都与词

汇能力描述和词汇能力测量密切相关。例如：基于语料库研究成果的扩展式词项 (larger lexical item) 将词与词之间的词汇、语法、语义和语用搭配范式看作是语言意义的延伸单位；计算词型和词符的比率是为了考查学习者的词汇广度发展状况；考查词汇短语而非单个词则体现了“嵌入式、综合型和语境扩充型”的词汇测试理念，等等。

基于上述观念的更新，作者提出了怎样才算学懂一个词项的问题。为此他介绍了 Nation (1990) 在 Richards (1976) 的词汇知识框架基础上发展起来的词汇知识构成表 (见 P. 26)。该表从接受和产出两个角度阐述了词汇知识应涵盖形式、功能和意义三个层面，即掌握一个词不仅要掌握它的发音和拼写形式，还要关注其深度知识，包括语法范式、词汇搭配、使用频数、语境特征、语义功能和语义关联等²。

关于什么是词汇能力，作者阐述了 Chapelle (1994) 的三大能力构成框架：一是在语境中把握词汇使用的能力；二是词汇知识及其最基本的加工能力；三是词汇应用的元认知策略。第一种能力包括在不同年龄人群、不同语言背景以及语域的背景下语境对词汇意义的影响。第二种能力往往最受二语教师和研究者的关注，它包括词汇量、词汇的各层面特征、心理词汇以及在理解和产出词汇时采用的各种心理加工程序。第三种能力也就是第一章中提到的语言策略使用能力，它包括在真实交际中使用的诸如释义法、语码转换、使用上义词表述下义词或者寻求帮助等各种产出型策略，以及诸如查字典或语境猜词等接受型策略。

●● 第三章 词汇习得和使用的研究综述

本章有四个重点话题：(1) 系统性词汇学习；(2) 随机性词汇学习；(3) 从语境中推导词义；(4) 交际中的词

² Nation 在其词汇知识分类表 (2003 年版本) 中进一步将词汇知识归纳为词汇形式、意义和使用三大层面并举例说明。详见 Nation, P. 2003. *Materials for teaching vocabulary* [A] In Tomlinson, B. (ed.) *Developing Materials for Language Teaching* [M]. London: Continuum. P. 394. ——本文作者注

汇应用策略。作者首先对与各个话题相关的一批研究进行了综述，然后反思这些研究成果对词汇测评的启示。

对系统性词汇学习的研究主要关注传统的词汇学习方法，如记忆配有母语解释或翻译的二语单词表。重点研究哪些特征使得它们比较容易或难以记忆；究竟有没有比死记硬背更好的记忆方法；是否应鼓励语言学习者使用记忆技巧学习词汇等。这是比较传统的、关注接受型而非产出型的词汇习得研究，比较适合应用分离式、选择型和脱离语境的词汇测评模式。在测试的设计上，不仅可以考查词的基本意义，还可以考查衍生的词形、拼写、同义词、搭配词和语法结构。

随机性词汇学习是指在完成阅读或听力训练任务时无意识或者顺带地习得一些词汇。这种习得已经得到母语或二语词汇习得研究的验证。测试时要注意的问题包括：要确保期待学生随机习得的词确实是生词；要注意研究中测试时机的控制，例如：是阅读和听完后立即测试，还是隔一段时间再测试；更要小心诠释学生“不经意习得的生词”的质量，即这类习得能够掌握生词到何种程度。

从语境中推导词义是二语习得研究中很热门的话题。它要求学生在自然交际语境中理解不熟悉的词汇。已有的研究关注的是什么样的语境信息有助于读者推导词义；这种信息在未经加工过的语篇中是否存在；未经特殊训练的学习者在推导词义时表现如何；推导能力可否经过训练获得；成功的推导是否真的能实现词汇习得等。研究中要注意的问题有：要弄清楚词汇推导测试的目的；测试设计要根据测试的目的决定，设计时要考虑如何选择目标词并以何种方式呈现给考生；考生应采取何种答题方式才能反映推导过程。

对交际中词汇应用策略的研究不多，它关注的是学习者在交际中遇到措词表达困难时，如何有意识地使用某种策略来化解困境。在这里作者引述了 Tarone (1978) 的交际策略分类框架，包括采用释义法、逐字翻译或直接使用母语中的对应词、打手势或表演、向对方寻求帮助、或者采用回避甚至放弃表述等策略。其他学者也提出了不同的分

类方法。尽管研究中提到的有些概念较为宽泛,如“交际目的”、“意义结构”,但是多数研究关注的都是词汇问题,只有少量研究致力于评价策略训练的效果。词汇交际策略的评价有两种方法:一种是用嵌入式、综合型的方法衡量学习者在口语任务中的表现;另一种是分离式、选择型、且要求学习者使用至少一种交际策略的测试任务。

第四章 词汇评价研究的综述

本章首先指出在整个语言评价研究领域尚缺乏对词汇测试的足够关注,而二语词汇习得的研究者却开发出不少考查词汇知识和能力的模式和方法。因此,作者利用以往针对母语阅读和二语词汇习得的理论和研究成果来充实对二语词汇测试的研究,全面综述了词汇评价中的若干问题。具体包括:客观型词汇测试的分离式测评属性、选择类试题的目标词项和干扰词项的设定原则、提升词汇知识测试的信度、词汇广度和深度的测量、评价词汇知识的质量、词汇测试的命题信度、语境的作用以及完形填空类题型的各种命题方式等。

作者首先提出问题,然后引介一批相关的研究来探讨解决问题的办法。例如:在探讨影响词汇选择题 (multiple-choice) 难度的变量有哪些时,作者引述了三个相关案例。第一个案例探讨选择题干扰词项的有效性,结果发现最有效的干扰项是那些能嵌入题干的语境或者与目标词语义接近的词语,结论是干扰项过于影响选择题对词汇知识的测评。第二和第三个案例探讨目标词的什么特征影响词汇选择题的难度,结果发现最常见的有频率、音节数和抽象程度。最容易选对的目标项是那些较经常出现、音节较少和语义较具体的词语;目标项的选对率还与它们是否容易发音、是否常在课本中出现、在源文中是用汉字还是拼音书写等因素有很大关系。

在探讨另一类测试题型,即完形填空题 (cloze test) 时,

作者通过研究个案来呈现这类题型的多个变体,包括标准完形填空 (standard cloze)、合理删词完形填空 (rational cloze)、选择式完形填空 (multiple-choice cloze) 以及首字母完形填空 (C-test), 介绍了各自的测试设计、局限性等。然后归纳出这类填空型测试模式的多个视角:(1) 它是分离式、选择型、语境依赖型的词汇测试形式;(2) 它考查的不仅是那些掏空了的词, 考生还要了解整个语篇中没有被掏空的实词;(3) 它考查的不仅是词汇知识, 也包括词汇技能和推导能力;(4) 它不仅可以考查个体的单词, 还可以拓展为考查与单词有搭配关系的短语或习语。

●● 第五章 词汇测试: 四个经典案例分析

本章重点剖析四个经典案例以探讨词汇测试的设计和效度问题。作者比较详细地介绍了每个案例的设计背景、测试理念、设计特点、实施方式、发展沿革、局限性等, 然后评价其效度和应用价值。

案例 1 是新西兰学者 Nation 在 20 世纪 80 年代初设计的词汇等级测试 (VLT), 供教师在课堂上使用, 帮助他们开发合适的词汇教学方法。基于越是常用的词汇越容易先被习得的理念, VLT 首先根据使用频数将英语词汇分成五个等级: 基本词汇 2,000 是 1 级, 3,000 是 2 级, 5,000 是 3 级, 第 4 级是超越 5,000 之后的大学词汇, 第 5 级是 1 万。在每个级别的测试题中, 设计 36 个单词, 18 个定义, 每组 6 个单词, 3 个定义, 要求考生进行单词和定义的配对。这样既可以减少阅读量, 又可以减少猜中答案的几率。每个级别考查的单词随机产生, 因此, 可以考查考生对这个级别单词掌握的比例。其测试形式是要求考生在从各级词汇中随机抽样出的词表中作单词和定义的配对。其作用是测试考生的词汇量和探究词汇广度的发展规律。它具有明显的分离式和脱离语境的特点, 操作简易快捷, 有利于考查接受型的词汇知识 (如只需知道词义), 但却不易考查产出

型的词汇知识。所以这种测试近年来也有创新,首先是扩大了测试的题量,另外,在形式上除了单词和词义配对之外,还出现了填空题。填空题考查的仍然是相同的目标词,设计者的目的是考查学生能否在有局部控制的条件下输出词汇。然而,填空测试很可能只是另一种考查接受型知识而非产出型知识的方法。

案例 2 是欧洲学者 Meara (1987; 1988) 和他的同事设计的欧洲语言中心词汇量测试 (EVST), 用于测评欧洲各国多所语言学校不同语言背景的英语学习者的词汇量, 以测试结果为基础进行分班教学, 学习不同的课程。与 VLT 一样, 它也是基于词汇使用的不同频数设计, 但分出了更多的等级, 而且使用电脑实施测评: 考生面对屏幕上逐一出现的一批从各级词汇表中抽取的目标词 (同时还掺杂部分乌有词), 点击按钮 “Yes” 或 “No” 来表示自己认识或不认识这个词, 全程不超过 10 分钟, 而且测试完后立刻就能看到得分。其特点是比 VLT 更加简便快捷。而且, 在题目中掺杂部分乌有词能够抵消该模式对测试效度的不利影响, 即防止考生点击一些原本不认识的词以获得高分。对该模式的后续研究集中在如何改善它的设计上: 改善测试的形式, 尤其注重乌有词的作用, 如根据不同语言背景的学生设置乌有词作为干扰项; 妥善选择要测试的单词; 以及如何更客观地诠释分值等。

案例 3 是加拿大学者 Paribakht 和 Wesche (1997) 设计的词汇知识等级测试 (VKS)。它要求考生判断自己对目标词的掌握符合以下五种描述中的哪一种: 没见过该词; 曾见过但不知道其意思; 曾见过而且觉得它的意思是 _____ (写出同义词或译义); 知道它的意思是 _____ (写出同义词或译义); 能够用该词造句 _____ (写出句子)。各层的分值从 1 排到 5, 第 3 到第 5 层只有在意思、拼写和用法完全写对的时候才能得到全分。1 到 5 层逐层推进, 由浅入深地考查考生对词汇掌握的程度。第 5 层旨在考查深度的知识产出。该模式的局限包括: 缺少二语词汇习得理论的支持; 不能同时考查一个词的多个词义; 用目标词造句的方式不一定能考查对单词意思的掌握; 一个固定的等级表不能够合理

地呈现人们对词汇的不同类型知识的掌握，可能需要多个等级表。

案例 4 是美国托福 (TOEFL) 考试中的词汇测试部分。托福考试依靠复杂的数据分析和测试技术保证测试质量和高效地测试大量的考生。作者对其中词汇部分的测试形式进行了较长时间的研究，结果发现 20 世纪 60 年代到 70 年代中期的模式是纯粹分离式和脱离语境的选择题设计，为此托福考试受到不少批评；自 70 年代中期起到 1995 年，托福考试将单词放在句子中考查，有了一定的语境；1995 年的考题改革干脆取消了独立的词汇测试题，将词汇放在阅读理解部分考查；从 1998 年起，托福考试的测试手段从纸笔考试发展为计算机考试，词汇测试的形式是为一段语篇中的某个单词从提供的选项中找出近义词，因此，并不代表向着更加依赖语境的测试方法发展。作者认为 1995 版的测试形式最能要求考生利用考题中的语境，但是这种潜力没得到充分发掘，而现在的词汇题是向着不依赖语境的词汇评价方式倒退了一步。

●●第六章 分离式词汇测试设计

本章首先论述影响分离式词汇测试设计的四个因素，然后辅以两个实例说明。

第一是明确测试的目的。总体来说，词汇测试有三个目的：研究；为教师和出题人提供决策支持；为项目提供决策支持。研究者研究词汇测试是为了更好地了解词汇知识的本质以及词汇习得的过程。教师和出题人则利用测试来分班、诊断问题、衡量进步和成绩、测量词汇熟练程度。在项目评估上，测试用来判断词汇学习项目的有效性或评估语言教育项目。

第二是界定测试要考查的词汇知识和能力。有两种界定方法，一种是基于教学大纲，另一种是基于理论。在课程学习的背景下评价词汇，可以基于教学大纲进行定义，这样可以根据课堂学习目标确定要评价的词项和词汇能力。

对以研究为目的的水平测试,要以理论为基础定义测试要考查的词汇能力。因为词汇知识、词汇习得和使用的理论和框架很多,所以定义起来比较困难。作者着重阐述了接受型和产出型知识这一对概念,因为它们在词汇评价中应用得很多,但是二者的区分仍不清楚,需要澄清。作者引用其他术语解释了二者的区别,首先是认识 (recognition) 和回忆 (recall),然后是理解 (comprehension) 和使用 (use)。作者认为接受和产出这两个术语太宽泛,在词汇评价时要弄清每个指的是什么样的学习者能力,才能更好地设计测试任务。

第三是测试内容的特点。选择型词汇测试设计的一个重要问题就是如何选择目标词。在目标词设置时可以参照“高频词”、“低频词”、“专业词”或“半专业词”的划分;而呈现目标词的方式则包括孤立呈现还是在语境中呈现,并不是所有词汇测试都要语境化,而是要根据不同的测试目的考虑语境的作用。

第四是答题的特点。在某些测试中,可以让考生评价自己的词汇知识,如前述的 EVST 测试就采用自我报告 (self-report) 的形式。在词汇评价中,自我报告有一定的作用,尤其是在测试词汇量时。然而,大多数测试都需要证据证明考生确实具备测试要考查的知识和能力,所以会要求考生为自认为已经懂得的单词写出定义或使用其造句,作为验证。测试设计还要考虑测试使用的语言,包括命题使用的语言以及答题所用的语言。单语测试形式只采用目标语一种语言,大多数词汇测试都采用这种形式;双语测试则既使用目标语又使用考生母语。

接下来作者列举了两个实例以说明上述因素在具体词汇测试中的体现。第一个例子是课堂进步测试。测试目的通常是考查学生在词汇学习上的进步,更是为了刺激他们有规律地学习词汇。这种测试很显然是根据教学大纲明确要考查的词汇能力,如根据上课计划确定是评价词汇能力的不同方面还是只有词义。在作者自己设计的词汇进步测试中,目标词通常选自课本中的某个单元或者专门的词表。在这种测试中,选择测试形式非常重要,因为考题需要容