

● 杨惠中 朱正才 方绪军 著

国家哲学社会科学规划项目



中国语言能力等级共同量表研究： 理论、方法与实证研究

The Development of
National Common Framework of Language Proficiency Scales
—Principles, Approaches, and Validation Study



上海外语教育出版社

外教社 SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS

www.sflep.com

国家哲学社会科学

● 杨惠中 朱正才 方绪军 著



中国语言能力等级共同量表研究： 理论、方法与实证研究

The Development of
National Common Framework of Language Proficiency Scales
—Principles, Approaches, and Validation Study

图书在版编目 (CIP) 数据

中国语言能力等级共同量表研究: 理论、方法与实证研究 / 杨惠中, 朱正才, 方绪军著. —上海: 上海外语教育出版社, 2012

国家哲学社会科学基金项目

ISBN 978-7-5446-2874-7

I. ①中… II. ①杨… ②朱… ③方… III. ①外语—语言能力—等级—研究—中国 IV. ①H3

中国版本图书馆 CIP 数据核字 (2012) 第 188001 号

出版发行: 上海外语教育出版社

(上海外国语大学内) 邮编: 200083

电 话: 021-65425300 (总机)

电子邮箱: bookinfo@sflap.com.cn

网 址: <http://www.sflap.com.cn> <http://www.sflap.com>

责任编辑: 杨 帆

印 刷: 上海华业装璜印刷厂有限公司

开 本: 787×965 1/16 印张 11 字数 180 千字

版 次: 2012 年 10 月第 1 版 2012 年 10 月第 1 次印刷

书 号: ISBN 978-7-5446-2874-7 / H · 1399

定 价: 28.00 元

本版图书如有印装质量问题, 可向本社调换

序 言

桂诗春

改革开放以来，适应着强大的社会需要，英语教育和测试得到空前的发展；而在当前商业大潮冲击下，它们也无疑地成为炒作得最为火红的一个“产业”，一个牵动人心的社会链接。它们既有许多利益相关者，还有更多的对英语赋予了许多附加值的社会关心者。英语无非是一种外国语，我们学习外国语的根本目标是什么？这个根本目标反而被人们置之脑后，连一天到晚忙于英语教学、特别是忙于帮助学生对付英语测试的老师也往往回答不上来。他们也许会说，“根本目标不就是提高语言能力嘛！”可是语言能力代表什么？却往往又若隐若现，难以触摸。我十分高兴地看到杨惠中、朱正才、方绪军三位作者所做的全国语言能力等级共同量表研究，觉得它虽然不是鸿篇巨制，但确实是一个严肃的、认真的科学成品，在现在还有人正本清源，擘肌分理，对语言（主要是英语）教学和测试的能力量表这样的基础工程进行研究，实属难能可贵。

这项研究说它是一个基础工程，并不为过。它不仅牵涉到语言教学和测试的理论和语言能力的描述，还牵涉到许多制定量表的原则和方法以及对能力量表的实验性的检测。而且从更广义的角度看，这还牵涉到对语言性质和语言教学思想的一些根本的看法，就外语教学而言，1880年来欧洲所掀起的外语改革运动提出现代语言的教学不能再像古典语言教学那样，把语言作为一门知识来教学，而必须强调语言能力的培养。经历过两次世界大战，世界格局有了很大的变化，到了上一世纪的60年代，语言学家Chomsky提出语言能力和语言运用的两分法，而牛津哲学家Austin却写了《怎样用语言做事》这一本影响深远的书。

在他的影响下，英人Wilkins把Austin的哲学思考引进了语言教学，写出了《意念教学大纲》。这个大纲为当时欧洲委员会制定欧洲统一的语言力量表奠定了基础。这个量表最早叫做《欧洲成年人现代语言学习的单元/学分体系》(A European Unit/Credit System for Modern Language Learning by Adults)，适用于欧洲委员会所属国家里成年人彼此学习对方语言的一个统一的语言体系，它不可能以具体的语言作为统一的标准，而只能以意念作为制定标准。当然意念也不一定反映真实的语言需要，后来又改称为功能-意念，具体化为“能做”的指标。这不仅是制定大纲和量表的描述语，而且是对语言在交际活动中的一种新的哲学思考，牵涉语言教学思想的重大改变。所以说语言力量表的制定对语言教学和语言测试都具有举足轻重的意义。

当然语言力量表的制定同时也是一项巨大的工程，本书的几位作者在理论和实践上都做了许多有益的探索，起到了先行者的作用。目前的研究只是较多地对英语口语的能力进行观察和描述，对制定整体的、全面的语言力量表虽然也能起到举一反三的作用，但是也还有许多工作需要继续做，例如对书面语的能力要求也很值得研究，因为中国学生学习英语往往不顾语体的差别，他们写的东西实际上是“写话”。

另外还有一项值得外语教学领导部门深刻考虑的事情，语言教学和语言测试的语言力量表实际上是一份语言教育的“行动纲领”——一方面纲要需要继续完善；另一方面纲要是需要实行的。怎样让它在实践中得以检验和贯彻，才能真正地不辜负三位作者的努力，改变我国外语教学的面貌，这比空嚷“听说领先”要重要得多，艰苦得多。

前言

近年来关于建立我国语言能力等级共同量表的问题引起了国内学界的重视。

我国是语言教学大国，各种外语的教学和测试、各少数民族语言的教学和测试、汉语作为外语或第二语言的教学和测试，在制度、规模、等级、组织形式等方面纷繁复杂。但是我国目前尚没有统一的语言能力等级共同量表。

制定统一的“语言能力等级共同量表”，也就是对语言能力制定统一的规格，为我国语言教学、语言学习与语言测试提供共同参照的框架与标准，这样做将有助于避免各语言教学与测试机构在语言能力描述方面各行其是现状，有利于教学，有利于课程的组织和衔接，避免课程的重复，做到外语教学大、中、小学“一条龙”，也有利于考试取得有科学定义的语言能力资格认证，便于用人单位对人才的合理使用。如果各种考试相互之间缺乏统一的能力描写标准，甚至连等级的划分和等级的数量也各有不同，用户将莫衷一是，不便于准确利用考试结果。

制定统一的语言能力等级量表，涉及测试项目的开发，涉及资格认证，也涉及教学过程组织、教材开发，可以说涉及与语言教学、语言测试有关的一切方面，是语言测试与语言教学领域的基础研究，在当前有着迫切的社会需求。

制定语言能力等级共同量表工程浩大，考虑到我国目前语言教学和语言测试在组织、机构、层次等方面种类繁多，规模庞大，实际情况复杂，要毕其功于一役是不现实的。一条可行的途径是先从局部做起，分阶段逐步实现制定共同量表的目标。比如，当前可以先着手制定我国英语能力等级共同量表，这既十分迫切，也具有现实可行性，并且在研究原则和方法上可以作为一种先行的尝试。

本研究得到了国家社会科学基金的资助,使我们有可能对建立语言能力等级共同量表所涉及的理论、原则、方法等进行深入、全面的研究。我们的研究集中在英语口语能力的描述和分级上。但是,语言测试研究在本质上是一种实证研究,对于所提出的原则和方法,只有经过较大规模的实证研究,才能验证其科学性和可行性。为此,前后有 14 所高校的 185 位资深外语教师参加了实证研究,为本项研究取得成功创造了条件。这些资深教师工作十分认真仔细,但是由于科研经费极其有限,根本无法酬报他们付出的辛勤劳动,这些教师参加这一项工作完全是出于为我国外语教学服务的奉献精神,令人敬佩,谨此表示由衷的感谢。

建立我国语言能力等级共同量表还是一项新的尝试,我们所做的仅仅是一些探索,由于各种条件的限制,本报告中提出的理论、原则、方法可能有许多不足,欢迎从事语言测试研究的同仁批评指正。

鉴于制定我国语言能力等级共同量表是一项基础性研究工作,不是几个人、几个部门能够完成的,希望这一项重要的工作能够引起我国有关部门的重视,大力投入人力物力开展研究,造福我国广大学子。

作 者

2011 年 10 月

目 录

序言	I
前言	III
 第一章 制定语言能力等级共同量表的意义	1
1.1 语言能力等级量表	1
1.1.1 语言能力等级量表的概念	1
1.1.2 语言能力等级量表的基本要件	3
1.2 制定语言能力等级共同量表的国际背景	9
1.3 制定中国语言能力等级共同量表的必要性	11
 第二章 语言能力理论及语言能力描述	16
2.1 语言能力的概念	16
2.1.1 概说	16
2.1.2 结构主义的语言能力观	17
2.1.3 Chomsky 的语言能力观	19
2.1.4 交际语言能力观	20
2.2 语言能力的维度	21
2.2.1 语言能力单维说	22
2.2.2 语言能力二维模型	23
2.2.3 交际语言能力模型	24
2.3 语言能力描述	29
2.3.1 语言能力、语言技能与语言水平	29
2.3.2 语言技能的描述	30
2.3.3 语言能力等级划分及评价的权重	40
2.4 小结	42

第三章 语言能力的“能做”描述:以 CEFR 为例 44

- 3.1 语言能力“能做”描述的背景 44
- 3.2 CEFR 描述语言能力的基本框架 47
- 3.3 语言能力“能做”描述语的量表化 49
 - 3.3.1 对描述语的基本要求 49
 - 3.3.2 语言能力描述语量表化的方法 50
- 3.4 CEFR 对语言能力描述的实用性考量 54
 - 3.4.1 面向不同人群的语言能力等级量表 54
 - 3.4.2 各等级描述的详略 55
- 3.5 CEFR 对语言能力的描述 57
 - 3.5.1 总体描述 57
 - 3.5.2 具体描述 59
- 3.6 CEFR 与 ALTE 量表 68
- 3.7 小结 72

第四章 制定中国语言能力等级共同量表的原则与

方法 74

- 4.1 制定中国语言能力等级共同量表的目标 74
- 4.2 制定中国语言能力等级共同量表的原则 76
 - 4.2.1 以交际语言能力理论为理论基础 77
 - 4.2.2 对听、说、读、写能力进行分别描述 77
 - 4.2.3 对语言能力进行“能做”描述 79
 - 4.2.4 语言能力等级描述详简有别 80
 - 4.2.5 便于理解和使用 81
- 4.3 制定中国语言能力等级共同量表的方法 82
 - 4.3.1 确定语言能力描述的参数项目 82
 - 4.3.2 建立描述语库 83
 - 4.3.3 语言能力描述语的量表化 84
- 4.4 建立语言能力等级共同量表与具体的语言测试项目的
关联 87
- 4.5 小结 88

第五章 英语学习者口语能力描述	90
5.1 口语能力结构及成分	90
5.1.1 口语交际的特点	90
5.1.2 口语能力的构成	93
5.1.3 口语能力描述的参数层级	101
5.2 口语能力描述实践中的参数设置	103
5.2.1 不同用途的口语能力描述	103
5.2.2 口语能力描述实践中的参数设置	105
5.3 建立英语口语能力描述语库	110
5.3.1 口语能力描述框架及基本参数	110
5.3.2 建立口语能力描述语库	113
5.4 描述语使用情况调查检验	120
5.5 小结	123
 第六章 英语口语能力描述语因子分析及等级划分	
实证研究	125
6.1 口语能力描述语调查分析基本路线	125
6.1.1 口语能力描述语使用情况调查设计	125
6.1.2 描述语使用情况调查数据分析处理基本思路	129
6.2 研究结果	131
6.2.1 三个公因子	131
6.2.2 聚类分析结果	132
6.3 讨论	133
6.3.1 因子一分析	133
6.3.2 因子二分析	135
6.3.3 因子三分析	137
6.3.4 因子分析与口语能力描述的方面	139
6.3.5 描述语所描述能力的等级划分	141
6.4 小结	144
 第七章 结语	146

参考文献	151
附录一 本项目收集的口语教学和测试大纲或评分标准 (17 种)	157
附录二 中国大学生英语口语能力描述语调查表	159
附录三 中国大学生英语口语能力描述语调查实验样本	165

第一章

制定语言能力等级共同量表的意义

1.1 语言能力等级量表

1.1.1 语言能力等级量表的概念

人们使用语言（包括母语和外语）进行交际的能力水平是有高低之分的。为了有效安排不同水平阶段的语言教学，为了测试语言学习者的语言能力水平，我们需要对人们的语言能力水平进行描述或评价，因此，就需要一套描述或评价的依据或工具。语言能力等级量表就是描述或评价人的语言能力水平的工具。

衡量人的语言能力水平高低的活动属于心理测量。人们一般把各种用于测量的量表分为四类，即称名量表（Nominal scale）、顺序量表（Ordinal scale）、等距量表（Interval scale）和比率量表（Ratio scale）（Bachman 1990）。跟语言教学和测试关系比较密切的是称名量表、顺序量表和等距量表三种。Brown（2005）用“连续量表（Continuous scales）”取代了“等距量表（Interval scale）”的概念，他提到跟语言测试有关的三种量表是称名量表、顺序量表和连续量表。就一般的记分体制（如“百分制”、“正态分制”）而言，称“连续量表”可能比“等距量表”更符合实际。因为语言测试的分数之间的距离很难像长度或重量的度量一样，做到相同的度量跨距之间是相等的。如，表示长度的95厘米与90厘米之间相差5厘米，85厘米与80厘米之间也相差5厘米，这两个5厘米是等距的。但作为语言测试的分数95分与90分之间相差5分，85分与80分之

间也相差 5 分，这两个 5 分的跨距是不是相等，就很难确定。从这个意义上说，用“连续量表”的名称来指称语言测试中的一种记分制度要准确一些。但目前人们习惯上还是用“等距量表”的名称，并规定用这种量表报告的分数每个单位的跨距是相等的，考生在一次测试中不同项目上的得分可以进行加减从而得到总分。

对语言能力的描述和评价通常使用顺序量表和等距量表。这两种量表都可以用来对人们不同的语言能力水平进行排序。如，假定采用顺序量表将语言能力分为六个等级，定义“一级”为最低，那么，这六个等级的语言能力水平就可以排序为（“>”表示“高于”）：

六级 > 五级 > 四级 > 三级 > 二级 > 一级

若采用等距量表（比如，用百分制）来描述语言能力，每一分之间的跨距是相等的，分数越高表示语言能力越强，那么：

90 分 > 80 分 > 70 分 > 60 分 > 50 分 > 40 分 ……

顺序量表和等距量表之间的重要区别在于，顺序量表的等级之间的跨距（即等级与等级之间的跨度）不一定相等，而等距量表的每个分数单位之间的距离假定是相等的，即具有等距（equal intervals）的属性。Bachman（1990）将顺序量表与等距量表作了对比，大意可以图示为（图 1-1）：

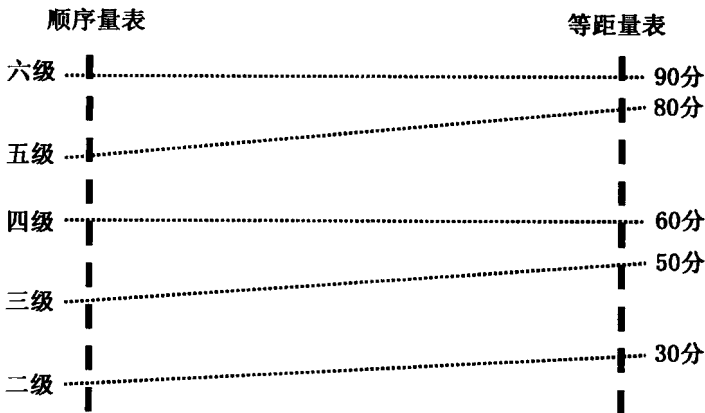


图 1-1 顺序量表与等距量表对比

从图示可以看出，与等距量表上的分数相比，假定顺序量表上的二、三、四、五和六级所表示的语言能力水平分别相当于等距量表上的 30 分、50 分、60 分、80 分和 90 分所表示的水平，那么，顺序量表上的

“三级”和“四级”之间的跨距是 10 分，而“四级”和“五级”之间的跨距则是 20 分，两个相邻等级之间的跨距不相等。从理论上说，理想的语言能力等级量表相邻等级之间的跨距应该是相等的，这样，便于理解和使用。比如，把英语能力水平分为六个等级或八个等级，最好各等级之间的距离是相等的，这样便于安排不同阶段的教学，便于设计适用于不同阶段的语言测试，也便于一般用户理解和使用。

1.1.2 语言能力等级量表的基本要件

制作用于测量物理量的工具，比如用于测量温度的温度计、用于测量长度的尺子，通常需要三个基本条件：设定零点，划分刻度，设定量表长度。制定语言能力等级量表也一样，需要确定语言能力等级量表的零点、刻度和长度，也就是说，零点、刻度和长度是语言能力等级量表的基本要件。

(1) 设定零点（或参照点）

设定量表上的零点（或参照点）是计量的最基本的条件。量表上的零点就是量表用于测量某种量度的起点。比如，把水在一般情况下结冰时的温度定义为零度，以此为基准，才能制造出温度计。温度计上的零度是相对的，“零度”并不意味着“没有温度”。语言力量表上的零点也是相对的，即语言力量表上的零点并不表示“没有语言能力”，它只是表示描述或评价语言能力的基准、起点。就英语能力等级量表而言，我们可以把学生刚开始学英语时的英语能力定为零点，也可以将学生进入高中时的英语能力定为零点，基于这两个不同的零点可以构造出不同的等级量表，而这两个量表上的相应等级所表示的语言能力水平很可能是不同的，比如，在前一个量表上的二级不一定等于后一个量表上的二级，就像摄氏零度（水结冰时的温度）不等于华氏零度（水的冰点温度是华氏 32 度，华氏零度比摄氏零度所表示的温度更低）一样。

在语言能力等级量表上，零点设定在哪里从本质上说可以是任意的，但由于零点在量表上是确定其他较高和较低等级分界的参照基准，因此必须对零点所表示的语言能力水平进行精细、清楚的描述，以便为确定并描述其上或其下等级提供清楚的参照。

(2) 划分刻度

语言能力等级量表要对高低不同的语言能力水平进行描述,就要确定量表的刻度,即确定语言能力等级的跨度及等级之间的分界,就像要用尺子度量长度,就必须在尺子上标定刻度一样。比如,在尺子上标明刻度时要决定是用“尺”、“寸”等为单位,还是用“米”、“厘米”等为单位。如果两个量表上的刻度单位长短不同,那么用这两个量表测得的数量是很难直接比较高低、大小的。比如,就像“一米”不等于“一尺”一样,大学英语考试(CET)的“四级”与英语专业考试(TEM)的“四级”也是不同的。

定义语言能力等级量表上的刻度需要解决两个问题:一是决定刻度的大小;二是如何使等级之间的跨度等距。

语言力量表上刻度的大小跟描述或测量的精度和操作的难度有关。首先,刻度的大小跟描述或测量的精度有关。用于描述或测量同一人群语言能力的量表,量表上的刻度之间的距离(即跨度)越大,描述或测量的结果就越粗;反之,描述或测量的结果则越细。比如,就构造中国英语学习者的英语能力等级量表而言,在量表上划分出八个等级自然要比仅仅划分初、中、高三个等级精细得多。

其次,刻度的大小跟构造量表和运用量表进行测量或评价时操作的难度有关。为了使描述或测量的结果达到更高的精度,量表上的刻度就需要更细一些,相应地,就需要将语言能力水平划分为更多的等级。但对语言能力等级的区分并不像在尺子上标明“厘米”的刻度那样简单。对语言能力等级区分越细,区分时的操作难度就越大。将语言能力区分为八个等级比仅仅区分初、中、高三个等级要难得多。

从理论上说,在语言能力等级量表上标记刻度、设定刻度之间跨度的大小,都可以是任意的。假定需要描述或评价语言能力水平的人群的语言能力水平在“0—100分”之间(用百分制表示),那么,在这“0—100分”之间划分出多少个等级,各等级包含哪些分数段,可以是任意的。比如,我们可以根据需要对这“0—100分”进行如下划分,从而形成六个量表(表1-1):

表 1-1 刻度不同的量表

100 分	量表一	量表二	量表三	量表四	量表五	量表六	
0 - 10	一级	一级	一级	一级	一级	一级	
11 - 20	二级		二级			二级	二级
21 - 30	三级		三级				
31 - 40	四级	二级	四级	二级	二级	三级	
41 - 50	五级	三级	五级	三级	三级		
51 - 60	六级	四级	六级	四级	四级		
61 - 70	七级	五级	七级	五级	五级	四级	
71 - 80	八级	六级	八级			六级	五级
81 - 90	九级	七级		六级			
91 - 100	十级	八级					

在上表中,如果要求刻度间的跨度是等距的,那么“量表一”和“量表六”是等距量表,其他的量表均为顺序量表。“量表二”和“量表三”虽然都分了八个等级,但相应的等级所表示的语言能力水平并不相同,比如“量表二”中的四级与“量表三”中的四级所表示的语言能力水平是不同的。“量表四”和“量表五”对低端和中段的语言能力等级划分是相同的,但对高端语言能力等级的处理不同:“量表四”把“四级”以上的语言能力再分为两段,而“量表五”对“四级”以上的语言能力则不作区分。“量表五”和“量表六”虽然都分了五级,但“量表五”对中段语言能力区分较细,对低端和高端语言能力则未加细分,而“量表六”对各段语言能力进行了等距切分。就以上六个量表抽象地看,很难判断它们各自的优劣,因为一个等级量表设立多少等级、各等级之间的跨度多大,这是可以根据实际需要来划定的:如果需要对各等级语言能力的全距都进行比较具体的描述或评价,那么可以采用与

“量表一”和“量表六”类似的量表（“量表一”比“量表六”的精度高）；如果需要对高端的语言能力进行更具体的描述，可以采用类似“量表二”的量表；如果需要对低端的语言能力进行更具体的描述，可以采用类似“量表三”的量表；如果只要对中段的语言能力进行更具体的描述，可以采用类似“量表四”或“量表五”的量表。

因此，光看以上几种量表对语言能力水平的等级划分，很难孤立地判断其优劣。在制定具体的语言能力等级量表时，可以根据对语言能力水平描述或评价精度的需要，选择等级划分合适的量表。

从便于使用和理解的角度考虑，语言力量表上的等级跨度最好是等距的。但如何使不同等级之间的跨度做到等距，则需要精心设计、严密论证。North（2000）根据统计结果，采用每个等级大约跨一个 logit 的方法^①，使划分出的每个等级基本等距，这种方法值得参考。

（3）设定量表长度

语言能力等级量表的长度跟描述或评价语言能力的高低范围有关。如果把描述或评价对象的语言能力水平看成是从“极低”到“极高”的线性连续体，对这个连续体进行全面的整体描述，就会得到一个最长的语言力量表。但构造这样的大而全的量表并不实用，因为语言能力极低和极高的人都是少数，对极低和极高的语言能力水平进行描述也没有多大的实用价值。既然不必对人们高低不同的语言能力水平连续体进行全面描述或测量，那么，在构造量表时如何确定语言能力“低点”和“高点”呢？实际上，这个“低点”和“高点”是可以根据实际需要任意设定的。比如，在“0—100 分”范围内，可以将“0—10 分”的语言能力水平定为“低点”，将“91—100 分”的语言能力水平定为“高点”（如表 1-2 中的“量表七”所示）；也可以将“31—40 分”的语言能力水平定为“低点”，将“91—100 分”的语言能力水平定为“高点”（如表 1-2 中的“量表八”所示）；还可以将“0—10 分”的语言能力水平定为“低点”，将“31—60 分”的语言能力水平定为“高点”（如表 1-2 中的“量表九”所示）；甚至可以将“31—50 分”的语言能力水平定为“低点”，将“61—70 分”的语言能力水平定为“高点”（如表 1-2 中的“量表十”所示）。也就是说，语言能力等级量表的长短也是可以根据需

① 实际上是中间略窄（0.97/0.98 logits），两头略宽（1.10 logits）。