

# 中国英语能力等级量表的“口语量表” 制定原则和方法\*

金 艳<sup>1</sup> 揭 薇<sup>2</sup>

**提要:** 本文报告了中国英语能力等级量表的“口语量表”制定原则和方法。首先,作者简要回顾了我国英语口语教学和测试现状,介绍了量表研制服务于教学和测试的目标以及基于实证研究的基本原则。其次,作者界定了交际语言能力视角下的口语能力构念,描述了口语量表的框架结构,即从口头交际活动、口头交际策略和口语文本特征三个维度描述口语交际能力。最后,作者阐述了基于实证研究的口语量表研制和效度论证的具体步骤与方法,并指出了口语量表研制和使用过程中有待探索的重要问题。

**关键词:** 中国英语能力等级量表; 口语量表; 量表制定原则; 量表制定方法

**Abstract:** This article reports the principles and methods of developing the Speaking Scale of the China's Standards of English (CSE-SS). First, based on a brief review of the status quo of oral English teaching and assessment, the authors set the goal of promoting oral English teaching, learning and assessment for the CSE-SS, and stressed the importance of conducting empirical investigations in developing the CSE-SS. Second, the authors defined the construct of oral English proficiency by drawing on the theory of communicative language competence. The authors then presented the overall structure of the CSE-SS framework, in which oral English proficiency is described by three categories of can-do descriptors: oral English communication activities, oral English communication strategies and typical linguistic features. Finally, the authors elaborated on the procedures and methods of constructing and validating the CSE-SS, and pointed out the future directions of research for the development and use of the CSE-SS.

**Key words:** China's Standards of English; Speaking Scale; principles of language scale construction; methods of language scale construction

中图分类号: H319 文献标识码: A 文章编号: 1004-5112(2017)02-0010-10

## 1. 口语量表项目背景

《欧洲语言共同参考框架》(以下简称《欧框》)(Council of Europe 2001)自 20 世纪 90 年代颁布以来,深刻影响着欧洲国家的语言教育政策,成为制定教学大纲、设置课程、编写教材、研发考试等的重要参考(Figueras 2012),是现代欧洲语言教育和学术界最具影响力的语言能力框架(Alderson 2005)。随着《欧框》在世界各地的广泛运用,我国的语言政策制定者、语言教师和测试工作者越来越深刻地认识到科学、统一、规范的语言能力量表对划分和描述语言能力等级的重要性(戴炜栋 2001;杨惠中,桂诗春 2007)。我国是英语教育大国,全国统一的英语能力等级量表将有助于系统、准确地描述英语学习者能力,为英语教学目标定位、教材开发、教学评价和测试以及大规模考试等提供指导(Jin *et al.* 2017;韩宝成 2006;方绪军等 2008)。

\* 本文得到 2015 年教育部哲学社会科学研究重大课题攻关项目“中国英语能力等级量表建设研究”(编号 15JZD049)的资助,口语量表建设为该项目的子项目。作者在此向项目组的邹绍艳、张晓艺、王隽、王华、张琳、刘黎岗、高森、程欣、周燕、严华等成员一并致谢。

英语的各项技能中,口语能力是我国英语学习者的能力短板之一。在基础教育阶段,很多地区的学校缺少英语口语学习环境,师资匮乏,学生在课堂内外训练英语口语的机会有限。在高等教育阶段,许多高校没有单独开设口语课程,也未能有效考核学生的英语口语能力。基于中国学习者英语口语语料库的相关研究发现,大学生英语口语的词块使用特征、话语结构特征和会话管理策略特征等不够明显,会话策略运用与本族语者差异较大,互动表达的形式和策略比较单一,实施语用功能的复现词块使用严重缺失(卫乃兴 2007; 许家金,许宗瑞 2007)。

造成学生英语口语能力相对薄弱的原因之一是缺乏明确、系统的教学目标和科学、有效的测试手段。为此,教育部考试中心于2014年启动“中国英语能力等级量表”(China's Standards of English, CSE)项目(林蕙青 2015; 刘建达 2015; 杨惠中 2015)。本文以该项目的子项目“中国英语能力等级量表的口语量表建设”为例,阐述量表的制定原则和方法,并报告口语量表建设的进展。

根据项目计划(参见教育部考试中心 2014),口语量表建设分为三个阶段:1)前期准备,2)框架构建,3)描述语库建设。第一阶段,项目组研讨了我国英语口语教学和测试现状,明确项目建设的目标和量表制定的基本原则。第二阶段,项目组通过文献回顾,进一步厘清了口语能力构念,并据此确定了量表的框架结构,即描述语的参数体系。第三阶段,项目组开展实质性的描述语库建设和效度验证:采集、改写和撰写描述语,通过定性和定量研究验证描述语的分级和分类,不断完善口语量表。以下分别介绍各阶段口语量表研究的成果和进展。

## 2. 口语量表建设的目标和原则

### 2.1 口语量表建设的目标和意义

制定语言量表的主要任务是科学描述语言能力,区分不同的水平等级(North & Schneider 1998; 方绪军等 2008)。根据CSE项目的总体目标,口语组确定了口语量表建设的具体目标:基于我国英语口语教学和测试现状,对接国家发展的英语口语能力需求,依据口语能力构念理论建立口语能力描述语库,并在实证研究基础上划分等级,撰写具有区别性特征的口语交际能力等级描述,为各阶段的英语口语教学和测试提供参照标准。

我们认为,口语量表对教学的指导作用主要在于明确英语口语教学的目标。量表将有利于教学的组织和衔接,有助于教师把教学整体划分为含有明确目标的各个小阶段,通过适合各阶段的教学方法更有效地组织课堂教学。量表也将有利于教材开发和编写,提升教育系统的透明度。

其次,量表的实施将使英语口语测试的分数解释更加清晰,分数使用更加公平。统一量表的制定能为英语口语考试研发提供科学、系统、清晰的评价指标体系,有助于教师设计学习成果的评估方法和工具,提高口语能力测试和评价的质量,更好地掌握和了解学生的实际口语水平。考试机构也可通过考试与量表的衔接提升不同考试之间的可比性,增强考试的透明度,以便考试用户更好地理解和使用考试结果,教育主管部门对不同地区、不同学校的口语教学进行更为科学的评估,促进我国英语测评与国际标准接轨。

### 2.2 口语量表制定的原则

口语量表制定遵循三项基本原则:第一,立足中国国情,服务于中国的英语教学和测试;第二,制定便于师生理解、运用的等级描述;第三,以语言运用为导向,关注口语交际活动和显性文本特征,开展实证研究,验证量表效度。

中国是当前国际英语教育和评价领域关注的热点,原因之一是我国英语学习者人数众多,

英语学习需求量大,能为具有商业运营能力的教育产业或考试机构带来利润。但是,教学和考试关乎国家的教育主权,因此量表制定要充分考虑我国英语口语教学和测试的现状,要为我国广大英语学习者服务,为改进我国外语教学和测试服务。同时,我们要借鉴国外的先进理念和方法,争取逐步在国际外语教学和测试领域掌握更多话语权。只有我们的量表得到国际认可,我们的课程、考试与量表有效对接,我国自主设计的考试才可能走出国门,成为评价英语能力的国际标准。

其次,近三十年来我国在英语口语教学和测试方面积累了丰富的数据和经验,开发了许多教材和测试。以测试为例,目前我国有大学英语四、六级口语考试和英语专业四、八级口语考试等。量表建设应充分利用现有的课程标准、教学要求以及评分标准,借助各教育阶段专家的判断,广泛征求教师和学生意见,制定适合我国外语教学和测试的量表。最近的一项调研发现,教师和学生对《欧框》描述语的理解有一定困难(Jin *et al.* 2015)。因此,量表研究应致力于制定易于理解且便于运用的等级描述。当然,量表宣传及相关知识的普及工作也十分重要。

第三,量表采用“能做”描述语(*can-do descriptor/statement*),描述具体交际情景下的语言运用,而不是抽象的语言能力。这一量表形式在世界上很多国家的英语教学和测评体系中广为应用,影响深远。因此,口语量表建设以描述语库为基础,既关注口头交际活动和完成活动所需的交际策略,也关注口语文本特征,通过描述语分级和分类的问卷调查、专家判断等大量定量、定性研究,实现描述语的量表化,这是制定语言能力等级量表的核心任务。

### 3. 口语量表的框架结构

与听、读、写相比,口语活动具有交互性、动态性、即时性等独特性,并且口语表达经常体现出副语言特征和非语言特征(Halliday 1989; 杨惠中等 2012),因而界定和描述口语能力是一项极为复杂的工程。基于交际语言能力理论中的口语能力定义,项目组从口头交际活动、口头交际策略和口语文本特征三个方面搭建口语量表的框架。

#### 3.1 口语能力构念

目前,尽管学界对于英语口语发展的阶段性特征尚未达成共识(Fulcher 2003; Luoma 2004),但是口语能力的描述角度相对明确,即从认知或交际角度来描述。在认知视角下,口语表达被看作一个程序性的有机整体,是一个过程性、整体性概念(Bygate 1987)。交际视角则以口语能力可分性为前提,阐述口语能力构念的成分,其研究路径主要脱胎于交际语言能力模型(如 Hymes 1972; Canale & Swain 1980; Bachman 1990)。交际视角关注语言交际活动及完成活动所需的能力成分,因此更适合作为量表构建的理论基础。

根据 Bachman(1990)的交际语言能力理论,语言能力(*language ability*)包括语言使用者掌握的语言知识(*language knowledge*)和指导、监控语言交际的策略能力(*strategic competence*)。Bachman & Palmer(1996)提出了包括六大类语言知识和三大类元认知策略的语言能力成分。“技能”并不是语言能力的一部分,而是语言能力在具体活动中的运用,即技能是任务特征(*task characteristics*)与语言能力的组合;因此,与其把“说”界定为一种抽象技能,不如将其描述为相关语言能力在具有某些特征语言活动中的具体运用(Bachman & Palmer 1996: 76)。

交际语言能力理论认为策略能力是交际能力的核心成分,交际策略是为了弥补交际时的语言能力缺位、取得最佳交际效果而采用的特定行为。Bachman & Palmer(2010: 52)指出,语言使用者在语言交际过程中运用元认知策略、语言知识和话题知识制定交际目的的实现计划,并运用认知策略实施计划,完成交际任务。

### 3.2 口语量表描述框架

口语量表的描述语分类操作性框架设有三级参数。第一级参数为框架的三个主要维度,即口头交际活动、口头交际策略和口语文本特征。第二级参数将交际活动分为口头表达和口头交流,将交际策略分为规划、执行、评估和补救策略,将文本特征分为语言、语篇和语用特征。第三级参数是构建分量表的具体参数,共有 17 项分类(见图 1)。



图 1 口语量表描述框架

首先,抽象的语言能力主要表现在具体语言运用之中,因此口语量表细致地描述各类口头交际活动。活动分类参考交际功能分类,即根据口语交际功能对活动进行系统分类。在语言教育领域,体裁( genre)概念被广泛运用于语言文本的分类。相同体裁的文本具有类似交际目的和使用情景,并由此拥有约定俗成的语言特点( Swales 1990; Hyland 2004)。根据口语文本体裁对应的交际功能( Halliday 1978; Martin & Rose 2003),口语量表描述以下口头交际活动:口头叙述(如个人经历、历史事件)、口头描述(如自然场景、人物心理)、口头说明(如报告)、口头论述(如演讲)、口头指示(如指令、程序)、事务性交谈(如谈判)和非事务性交谈(如聊天)。

其次,策略能力是语言能力的重要成分。项目组回顾了交际策略的相关研究(如 Færch & Kasper 1983; Tarone 1983; Dörnyei & Scott 1995)结合我国学者对交际策略理论的梳理(如戴曼纯 1992; 戴炜栋,束定芳 1994; 刘乃美 2005),并参考《欧框》的策略分类,将口头交际策略分为规划策略、成就策略(如转述、近似表达、使用概括词、重组、造词)、回避策略(如回避或转移话题、放弃或减省表达)、拖延策略(如使用填补词、重复、停顿、迂回)、合作策略(如提问、询问意见、给予反馈、引述对方话语)、话轮转换、自我修正和解决问题策略(如求助、回应对方求助、请求澄清、回应对方澄清请求)。

最后,口语文本特征对教学和测试具有重要的现实意义。根据现有英语口语能力量表(如 ACTFL 量表、ALTE 量表)和考试评分量表(如大学英语四、六级口语考试评分标准等)对口语能力的描述,结合 Hymes(1972)的交际语境框架和 Bachman & Palmer(1996)的语言运用

框架,项目组从准确性、广度、流利度、连贯性、灵活性、互动性和得体性等 7 个维度描述英语口语文本特征。

#### 4. 口语量表的建设方法

为制定口语量表研发和效度论证方案,项目组总结了国内外语言力量表的制定经验,分析和比较了不同方法的优势与不足,提出了定性和定量研究相结合的量表研制方案。目前,口语量表建设已经完成的工作包括口语描述语预备库建设、描述语查重和筛选、描述语分级、工作坊研究。

##### 4.1 口语量表研发和效度论证方案

语言能力等级量表的制定方法主要有直觉法、定性法和定量法。直觉法是早期语言力量表研制的主要方法,是基于经验的量表制定方法,即借助课程或测试专家的经验制定量表,往往在量表效度验证研究中受到挑战。North & Schneider(1998)认为,基于专家经验和直觉或者借鉴已有量表制定的语言力量表,在高风险决策环境中会产生各种问题。Fulcher(1996)在总结 ACTFL 量表效度研究的基础上指出,如果量表在研制初期缺乏扎实的实证基础,在投入使用后就很难对其进行效度研究,量表描述的能力很可能与其假设或界定的构念完全背离。

加拿大的 CLB 量表(Canadian Language Benchmarks)制定使用了直觉法和定性法,通过专家判断和学生自评的方式收集数据,采用特征分析、归类分析或协商等定性分析方法。《欧框》的研制采用直觉法、定性研究和定量研究相结合的方法。North(2000)采集了大量语言能力描述语,建立了语言能力描述语库,并运用 Rasch 模型对描述语进行分级验证,建立了 50 余个多维度、多等级的语言能力分量表。根据口语量表制定原则(见 2.2 节),我们拟定了定性和定量研究相结合的量表制定和效度论证方案,具体如图 2 所示。

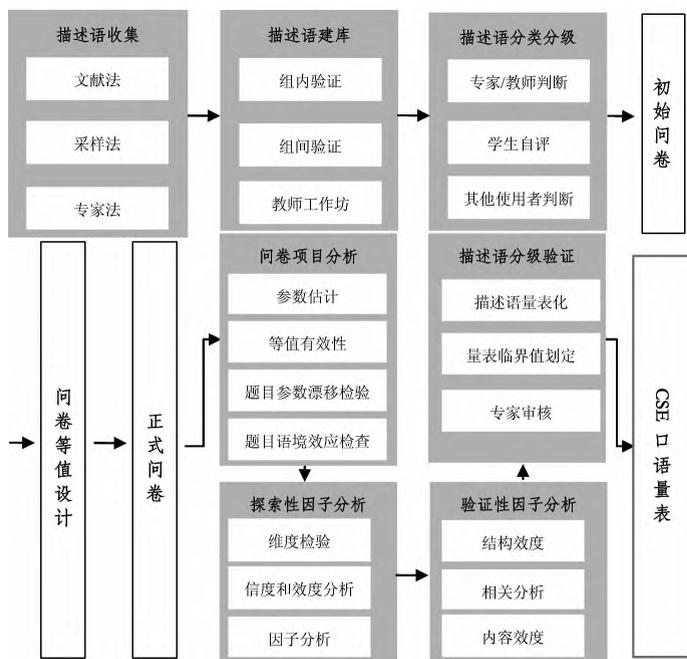


图 2 口语量表制定和效度论证方案

首先,项目组通过文献法、采样法和专家法收集各教育阶段的英语口语能力描述语,并对其进行筛选。文献法指通过查阅文献资料,对相关研究进行梳理总结,提炼相关能力的构念,形成量表的初始描述语。采样法指邀请教师、专家通过访谈编写或改写描述语。专家法即由相关领域专家基于使用者的意见和反馈,设计、撰写和修订描述语。这些方法在我国语言能力量表研究中已有一定实践基础(参见王淑花 2012;杨惠中等 2012)。

其次,项目组对入库的描述语进行初步分类和分级。分类指确定每条描述语描述的口语能力目标,并为每条描述语标记参数(参见图 1)。分级指将每条描述语归入相应的等级。在建库阶段,项目组进行了三次描述语分类和分级的初步验证。第一次为组内验证,即每条描述语由三名以上组员分类分级,组员如意见不一致则进行讨论或改写。第二次为组间验证,即与听力、阅读、写作等项目组进行交叉验证,保留一致性较高的描述语,修改或删除分歧较大的描述语。第三次验证采用工作坊形式,由教师审核每条描述语的文字表述及其分类分级,项目组在此基础上完成描述语验证库的初步建设工作。

量表大样本分级验证采用问卷方式,调查教师、学生和其他使用者对描述语分级的看法。项目组采用非等组锚测验设计( nonequivalent groups with anchor test) 形成若干套正式问卷,在全国各教育阶段的师生中开展广泛调研。正式问卷的数据分析分步进行。首先,通过多级计分项目反应理论模型估计项目参数,确定参数估计的方法,将不同级别的描述语转换到同一量表中。在垂直量表化过程中,由于锚题涉及的难度水平不同,需要考虑极端锚题带来的语境效应( Kolen & Brennan 2014)。链接问卷时还需实施题目参数漂移( Item Parameter Drift, IPD) 检验,测出 IPD 则需寻找原因并采取相应措施( Ye & Xin 2014)。其次,采用探索性因子分析验证量表的维度( dimensionality),确定最有效的因子结构,评估量表的信度和效度。接着,通过验证性因子分析验证量表的结构效度,检测因子与对应项目之间的关系是否与理论假设一致。最后,在描述语量表化阶段,项目组邀请各级别的专家参考数据分析结果,对描述语的分类和分级进行多轮审核,最终形成英语口语量表。

#### 4.2 口语描述语预备库建设

口语描述语库建设的方法主要包括已有量表提取、专家编写或改写、学生语料编写、教学目标任务分析等。描述语库建设参考国内外 118 份资料: 44 份综合量表(如 ACTFL 量表、ALTE 量表、《欧框》)、20 份评分量表(如大学英语四、六级口语考试评分标准)、42 份课程标准或教学大纲(如《高等学校英语专业英语教学大纲》、《大学英语课程教学要求(试行)》、New York State P-12 Common Core Learning Standards for English Language Arts & Literacy)、12 份考试大纲(如《高校英语专业四级口试大纲》)。通过整理、拆分、提取和录入,我们建立了包含 4 463 条中英文描述语的口语描述语预备库。为避免翻译引起的意义改变,这一阶段所有描述语均使用原材料的语言。

#### 4.3 描述语查重和筛选

口语描述语预备库中的 4 463 条描述语会不可避免地出现重复或相似现象。因此,在描述语分类和分级之前,先要进行描述语查重和筛选。培生标准和质量办公室颁布的描述语质量标准指出,“能做”描述语的开发与撰写须遵循两个维度的标准:质(能做什么)和量(能做到什么程度)( Pearson Standards and Quality Office 2014)。“能做”描述语的目标是描述运用语言能力能够完成的任务及任务的完成程度。为了实现这一目标,描述语应包含语言行为( performance)、语言行为的内在质量标准( criteria) 和支撑语言行为的其他外部条件( conditions) 三

项要素,如图 3(参见 Pearson Standards and Quality Office 2014)所示。

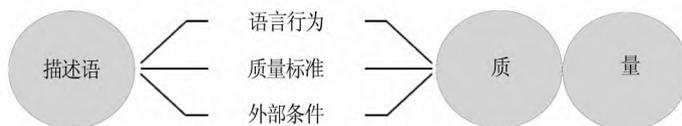


图 3 “能做”描述语的三要素

为开展描述语查重和筛选,项目组建立了口语描述语关键词表,然后根据描述语三要素模型,以关键词表为依据,利用统计软件 STATA 在描述语预备库中检索重复的表述。初步查重处理之后,描述语数量减至 2 487 条。经过逐条核查和多次讨论,项目组挑选、删除、合并重复和相似的描述语;再次查重和筛选这些描述语后,描述语数量精简至 908 条。在此基础上,项目组完成英语描述语的翻译、描述语语言风格和结构形式的统一。

#### 4.4 描述语分级

描述语入库时,需要逐条接受初步定级,即专家根据经验将描述语归入某个能力级别。基于我国英语教育阶段划分,同时参考现有语言力量表等级和课程标准,CSE 初步划分初、中、高三大等级共 9 个级别的英语能力水平。A 等(A1—A3)为初级,对应义务教育阶段的学习者英语水平;B 等(B1—B3)为中级,对应高中和高等教育阶段的英语学习者水平;C 等(C1—C3)为高级,对应高等教育英语专业学习者和英语精通使用者的水平(教育部考试中心 2014)。初步分类分级后,A1—A3 的描述语分别为 17、54、35 条,B1—B3 的描述语分别为 188、156、189 条,C1—C3 的描述语分别为 127、109、33 条,共计 908 条。各个级别的口头交际活动描述语示例如表 1 所示。

表 1 口头交际活动各级别描述语示例

类别	级别	描述语
描述	A1	能口头介绍自己的基本信息,如姓名、年龄和出生地等。
说明	A2	能在有提示的情况下,简单说明常见物品的用途。
指示	A3	能在完成小组合作任务时,用简短的口头指令分配角色任务。
论述	B1	能对同伴的口头阐述内容表达自己的看法。
叙述	B2	能详细讲述不同时期的个人经历,如求学或生活经历。
指示	B3	能就开展某种活动或计划的方法、步骤、程序等给予清晰、明确的口头指示。
论述	C1	能在演讲中根据需要采用合适的论证方法(如举例等)进行有力论证。
事务性交谈	C2	能在电话交谈中,就行程、价格等商务事宜进行有效的沟通和协商。
非事务性交谈	C3	能出色地主持或接受访谈,表达时充满自信、谈吐自如。

#### 4.5 工作坊研究

在“行为锚定”量表研制过程中,邀请量表使用者对描述语的分类和语义表述进行检验是一种常用方法(North 1993)。为此,本研究采用教师工作坊形式,了解教师对于量表和描述语

的理解程度,采集教师对描述语分类、分级的反馈信息和意见,并根据教师提出的评判依据,补充撰写新的描述语。在工作坊之前,项目组培训工作坊主持人、拟定主持提纲、组建讨论小组、明确讨论次序和时间。

项目组举办了 28 次小型教师工作坊,共有 48 名来自小学到研究生教育阶段且有 3 年以上教学经验的教师参加了工作坊。每次工作坊中,主持人首先采集教师个人信息,开展教师培训,培训材料包括口语量表描述框架和口语考试录像。随后,教师组成两人小组观看录像,评判录像中考生的表现,并对自己的评判给予详细解释(如“为什么考生甲的口语能力较考生乙强?”)。主持人录音并记录教师讨论过程。最后,主持人将初步划定级别的描述语发给教师,教师阅读每条描述语之后,根据教学经验判断各条描述语的分类是否合适、分级是否合理、表述是否清晰明确。

经过工作坊研究,908 条口语描述语删减至 823 条(见表 2)。

表 2 教师工作坊后的口语描述语数量统计

	A1	A2	A3	B1	B2	B3	C1	C2	C3	总数	
口头交际活动	(1) 口头叙述	3	4	3	2	6	4	5	2	0	29
	(2) 口头描述	1	5	6	10	4	6	2	5	0	39
	(3) 口头说明	1	3	3	16	4	5	10	5	0	47
	(4) 口头论述	0	3	2	3	4	8	7	3	1	31
	(5) 口头指示	0	3	2	4	2	3	0	0	0	14
	(6) 事务性交谈	0	7	7	25	10	5	5	3	1	63
	(7) 非事务性交谈	1	7	5	10	21	20	15	10	2	91
口头交际策略	(8) 规划策略	0	1	2	0	3	3	3	2	1	15
	(9) 执行策略	1	7	4	11	15	24	0	15	4	81
	(10) 评估和补救策略	0	5	5	4	6	13	4	2	1	40
口语文本特征	(11) 准确性	4	15	8	13	14	19	16	14	5	108
	(12) 广度	4	4	6	4	9	17	13	17	6	80
	(13) 流利度	1	3	3	2	3	9	6	9	5	41
	(14) 连贯性	0	3	4	2	3	12	4	5	1	34
	(15) 灵活性	0	2	4	2	4	3	6	2	2	25
	(16) 互动性	0	3	5	1	6	6	8	4	0	33
	(17) 得体性	3	3	4	2	4	15	12	6	3	52
合计	19	78	73	111	118	172	116	104	32	823	

口语描述语库初步建成之后,口语量表项目组将进入量表的效度论证阶段,进一步检验口语描述语的分类和分级,全面论证量表的效度。限于篇幅,这一部分的研究成果我们将另文报告。

## 5. 结语

中国英语能力等级量表制定是一项关乎我国英语教育改革发展的重点任务,制定统一的英语语言力量表必将有助于提高我国英语教学的质量和效率(林蕙青 2016)。本文主要从学术视角报告了口语量表建设的进展和成果。量表研制不仅学术性强,专业要求高,而且政策性强,社会影响力大(Alderson 2007; Jin *et al.* 2017)。《欧框》等语言力量表的开发和使用的经验告诉我们,语言力量表的合理使用比开发更具挑战性。如何在教学和测试中有效使用量表?如何避免过度或错误使用量表?这些都是中国英语能力等级量表建设须进一步探索的重要问题。

## 参 考 文 献

- [1] Alderson J C. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment* [M]. London: Continuum, 2005.
- [2] Alderson J C. The CEFR and the need for more research [J]. *The Modern Language Journal*, 2007, 91(4): 659-663.
- [3] Bachman L F. *Fundamental Considerations in Language Testing* [M]. Oxford: Oxford University Press, 1990.
- [4] Bachman L F & Palmer A S. *Language Testing in Practice: Designing and Developing Useful Language Tests* [M]. Oxford: Oxford University Press, 1996.
- [5] Bachman L F & Palmer A S. *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World* [M]. Oxford: Oxford University Press, 2010.
- [6] Bygate M. *Speaking* [M]. Oxford: Oxford University Press, 1987.
- [7] Canale M & Swain M. Theoretical bases of communicative approaches to second language teaching and testing [J]. *Applied Linguistics*, 1980, 1(1): 1-47.
- [8] Council of Europe. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* [M]. Cambridge: Cambridge University Press, 2001.
- [9] Dörnyei Z & Scott M L. Communication strategies: An empirical analysis with retrospection [A]. In Turley J S & Lusby K (eds.). *Selected Papers from the Proceedings of the 21st Annual Symposium of the Deseret Language and Linguistic Society* [C]. Provo, UT: Brigham Young University, 1995. 155-168.
- [10] Færch C & Kasper G. On identifying communication strategies in interlanguage production [A]. In Færch C & Kasper G (eds.). *Strategies in Interlanguage Communication* [C]. London: Longman, 1983. 210-238.
- [11] Figueras N. The impact of the CEFR [J]. *ELT Journal*, 2012, 66(4): 477-485.
- [12] Fulcher G. Invalidating validity claims for the ACTFL oral rating scale [J]. *System*, 1996, 24(2): 163-172.
- [13] Fulcher G. *Testing Second Language Speaking* [M]. London: Pearson Education, 2003.
- [14] Halliday M A K. *Language as Social Semiotic: The Social Interpretation of Language and Meaning* [M]. London: Edward Arnold, 1978.
- [15] Halliday M A K. *Spoken and Written Language* [M]. Oxford: Oxford University Press, 1989.
- [16] Hyland K. *Genre and Second Language Writing* [M]. Ann Arbor, MI: University of Michigan Press, 2004.
- [17] Hymes D H. On communicative competence [A]. In Pride J B & Holmes J (eds.). *Sociolinguistics* [C]. Harmondsworth: Penguin, 1972. 269-293.
- [18] Jin Y, Zou S & Zhang X. *What CEFR Level Descriptors Mean to College English Teachers in China* [R]. Toronto: The 37th Language Testing Research Colloquium, 2015.
- [19] Jin Y *et al.* Developing the China Standards of English: Challenges at macropolitical and micropolitical levels [J]. *Language Testing in Asia*, 2017, 7(1): 1-19.

- [20] Kolen M J & Brennan R L. *Test Equating ,Scaling ,and Linking: Methods and Practices* ( 3rd Ed.) [M]. New York: Springer-Verlag ,2014.
- [21] Luoma S. *Assessing Speaking* [M]. Cambridge: Cambridge University Press ,2004.
- [22] Martin J R & Rose D. *Working with Discourse: Meaning Beyond the Clause* [M]. London: Continuum ,2003.
- [23] North B. *Scales of Language Proficiency: A Survey of Some Existing Systems* [Z]. Strasbourg: Council of Europe ,1993.
- [24] North B. *The Development of a Common Framework Scale of Language Proficiency* [M]. New York: Peter Lang ,2000.
- [25] North B & Schneider G. Scaling descriptors for language proficiency scales [J]. *Language Testing* ,1998 ,15( 2) : 217-262.
- [26] Pearson Standards and Quality Office. *Writing Descriptors: Guidelines and Best Practice* [Z]. London: Pearson Publishing Ltd ,2014.
- [27] Swales J M. *Genre Analysis: English in Academic and Research Settings* [M]. Cambridge: Cambridge University Press ,1990.
- [28] Tarone E. On the variability of interlanguage systems [J]. *Applied Linguistics* ,1983 ,4( 2) : 142-163.
- [29] Ye M & Xin T. Effects of item parameter drift on vertical scaling with the nonequivalent groups with anchor test ( NEAT) design [J]. *Educational and Psychological Measurement* ,2014 ,74( 2) : 227-235.
- [30] 戴曼纯.第二语言习得者的交际策略初探[J].*外语界* ,1992 ( 3) : 8-12.
- [31] 戴炜栋.构建具有中国特色的英语教学“一条龙”体系[J].*外语教学与研究* 2001 ( 5) :322-327.
- [32] 戴炜栋,束定芳.外语交际中的交际策略研究及其理论意义——外语教学理论研究之三[J].*外国语* ,1994 ( 6) : 27-31.
- [33] 方绪军 杨惠中,朱正才.制定全国统一的语言能力等级量表的原则与方法[J].*现代外语* 2008 ( 4) : 380-387.
- [34] 韩宝成.国外语言力量表述评[J].*外语教学与研究* 2006 ( 6) : 443-450.
- [35] 教育部考试中心.中国英语能力等级量表工作手册[Z].北京:教育部考试中心 2014.
- [36] 林蕙青.深化考试招生制度改革 加强国家外语能力测评体系建设[J].*中国考试* 2015 ( 1) : 3-6.
- [37] 林蕙青.建立国家外语测评体系 提升国民语言能力[J].*中国考试* 2016 ( 12) : 3-4.
- [38] 刘建达.我国英语能力等级量表研制的基本思路[J].*中国考试* 2015 ( 1) : 7-11 ,15.
- [39] 刘乃美.交际策略研究对我国外语教学的启示[J].*外语界* 2005 ( 3) : 55-60.
- [40] 王淑花.中国学生英语理解力量表的构建及验证研究[M].北京:知识产权出版社 2012.
- [41] 卫乃兴.中国学生英语口语的短语学特征研究——COLSEC 语料库的词块证据分析[J].*现代外语* 2007 , ( 3) : 280-291.
- [42] 许家金,许宗瑞.中国大学生英语口语中的互动话语词块研究[J].*外语教学与研究* 2007 ( 6) : 437-443.
- [43] 杨惠中.关于我国外语能力测评体系建设的几点思考[J].*中国考试* 2015 ( 1) : 12-15.
- [44] 杨惠中 桂诗春.制定亚洲统一的英语语言能力等级量表[J].*中国外语* 2007 ( 2) : 34-37.
- [45] 杨惠中 朱正才,方绪军.中国语言能力等级共同量表研究: 理论、方法与实证研究[M].上海:上海外语教育出版社 2012.

作者单位: 1. 上海交通大学外国语学院,上海 200240  
2. 上海交通大学外国语学院,上海 200240; 上海对外经贸大学国际商务外语学院,上海 201620