

# 目 录

表目	XIII
图目	XV
<b>第一章 口译能力与量表评估</b>	<b>1</b>
1.1 口译能力	2
1.1.1 口译能力的定义	2
1.1.2 本科阶段的口译能力	4
1.1.3 口译能力评估	8
1.2 量表的起源和发展	9
1.3 量表构件	12
1.3.1 描述语	13
1.3.2 等级	15
1.3.3 典型活动	15
1.3.4 小结	16
1.4 量表的理论基础	17
1.4.1 自主学习理论	17
1.4.2 社会文化理论	20
1.5 小结	21
<b>第二章 基于量表的口译自我评估</b>	<b>22</b>

2.1 口译自我评估 .....	22
2.2 研究介绍 .....	24
2.2.1 案例背景 .....	24
2.2.2 量表 .....	25
2.2.3 自我评估过程 .....	26
2.2.4 数据收集和整理 .....	26
2.3 结果与讨论 .....	27
2.3.1 分指标对比 .....	27
2.3.2 分组对比 .....	29
2.4 结论和教学建议 .....	38
<b>第三章 基于量表的口译同伴评估与反馈 .....</b>	<b>40</b>
3.1 文献回顾 .....	40
3.1.1 同伴评估 .....	40
3.1.2 量表与量表培训 .....	43
3.2 口译量表培训一个案研究 .....	45
3.2.1 案例背景 .....	45
3.2.2 量表 .....	45
3.2.3 量表培训 .....	46
3.2.4 数据收集和整理 .....	48
3.2.5 评估活动分析框架：会话分析 .....	49
3.3 结果与讨论 .....	50
3.3.1 结构与功能 .....	50
3.3.2 功能与内容 .....	51
3.3.3 课堂话语与课后访谈 .....	53
3.4 结论和教学建议 .....	57
<b>第四章 基于量表的口译课堂双教师合作评估 .....</b>	<b>58</b>
4.1 文献回顾 .....	58

4.1.1 双教师合作历史 .....	58
4.1.2 双教师合作研究 .....	60
4.2 研究一 .....	61
4.2.1 研究背景 .....	61
4.2.2 量表 .....	62
4.2.3 评估过程 .....	63
4.2.4 数据收集和整理 .....	65
4.2.5 结果与讨论 .....	65
4.2.6 结论和教学建议 .....	69
4.3 研究二 .....	70
4.3.1 研究背景 .....	70
4.3.2 量表 .....	71
4.3.3 研究过程 .....	71
4.3.4 量表归纳能力和演绎能力 .....	72
4.3.5 研究结果 .....	73
4.3.6 小结 .....	76
<b>第五章 中国口译学习者课堂环境下的量表评估体系.....</b>	<b>77</b>
5.1 核心构念 .....	77
5.2 任务环境 .....	78
5.3 量表内容 .....	80
5.3.1 判断 .....	81
5.3.2 整合 .....	87
5.3.3 拓展 .....	90
5.3.4 小结 .....	92
5.4 量表实施 .....	93
5.4.1 评估主体 .....	93
5.4.2 组织方式 .....	94
5.4.3 模态选择 .....	96

5.5 量表评估体系构建 .....	96
结语.....	98
参考文献 .....	100
附录 1 自我评估口译材料 .....	111
附录 2 同伴评估口译材料 .....	113
附录 3 中外教师评估口译材料 .....	115

# 表 目

表 2.1	学生样本 .....	24
表 2.2	口译策略量表 .....	25
表 2.3	不同水平组评估指标的描述性统计 .....	27
表 2.4	量表文本整体特征分组比较 .....	30
表 2.5	量表文本例句对比 .....	30
表 2.6	量表文本分级对比 .....	31
表 2.7	学生 S1 自评过程 .....	35
表 2.8	学生 S3 自评过程 .....	36
表 3.1	口译策略量表 .....	46
表 3.2	量表培训两种活动对比 .....	48
表 3.3	话题单元下各话轮结构功能举例 .....	49
表 3.4	学生 E/F 评估与反馈片段 .....	53
表 4.1	合作授课中外教师背景介绍 .....	62
表 4.2	口译策略量表 .....	63
表 4.3	中外教师评估意见录音转写示例 .....	64
表 4.4	中外教师评估录音对比 .....	67
表 4.5	授课教师和学生情况（均使用化名） .....	70
表 4.6	口译策略量表 .....	71
表 4.7	量表归纳和演绎能力示例 .....	72
表 4.8	量表归纳能力双教师点评示例（逻辑策略） .....	73
表 4.9	量表归纳能力双教师点评示例（流畅策略） .....	74

表 4.10	量表演绎能力双教师点评示例（流畅策略）·····	75
表 5.1	口译能力总表（五级和六级）·····	79
表 5.2	口译能力自评表（五级和六级）·····	79
表 5.3	口译执行策略量表（节选）·····	83
表 5.4	口译理解和辨识策略（五级和六级）·····	88

# 图 目

图 1.1	口译能力阶段性图解 .....	5
图 2.1	量表使用的时间特征分组比较 .....	34
图 3.1	量表讨论分组展示 .....	48
图 3.2	第五周 1 班讨论话语的功能与讨论内容 .....	51
图 3.3	第五周 2 班讨论话语的功能与讨论内容 .....	52
图 4.1	教师评估标识示例 .....	64
图 4.2	评估标识中外教师对比 .....	65
图 5.1	量表评估体系 .....	97





# 第一章

## 口译能力与量表评估

### 本章要点：

- 描述和界定口译能力，根据本科阶段口译学习者的群体特征，提出“语言和内容基本准确、表达比较流畅”为该阶段口译能力的培养目标，确定《中国英语能力等级量表》口译子量表（五级和六级）为该阶段口译学习者量表评估的重点。
- 梳理量表的起源和发展，归纳中国量表与西方量表之间存在的社会现实差异：中国量表用于科举制度以求集权与公平，西方量表用于弱势群体以求保护个体。
- 分析量表构件，包括描述语、等级和典型活动。
- 述评量表助学的理论来源，包括自主学习视角下量表对学习者的情感调节和认知调节，社会文化视角下量表对学习者社交活动的指导等。
- 提出进一步拓展量表的应用场景，完善《中国英语能力等级量表》评估体系。

20 世纪上半叶，口译作为一门职业正式登上国际舞台，口译教学和口译研究得到了迅速发展。口译能力成为了教师、学者和译员共同关心的目标：对于口译教学而言，口译能力的研究直接影响着教学内容的组织和教学目标的制定，口译能力的培养也是开展口译教学的基础。对于口译研究而言，口译能力研究不仅拓展了口译过程、口译产品等应

用层面的研究，也深化了口译语言转换和认知思维等理论层面的研究，Schaffner & Adab (2000: xiv) 就认为，翻译理论的学者可以从翻译能力的角度加深对翻译程序的认识。此外，对于口译职业而言，口译能力的研究也有利于口译职业地位的确立，如 Pöchhacker (2004: 166) 所言，一门职业要想获得独立，就需要证明其具备一套复杂的职业知识和技能，且这些知识和技能只能通过专业的培训才能获得。因此，口译能力的研究是将口译理论、口译教学、口译职业三方面联系起来的重要纽带。

在我国，根据《普通高等学校本科外国语言文学类专业教学指南》(2020)，口译能力的培养和评估应该是英语专业必修的口译课程的中心环节。口译能力是一个连续体，具有等级进阶、成分丰富、视角多元的特点(苏伟，2011)，由于口译能力无法直接观察和测量，只能通过外在表征和内在机制推理而得，所以开发出一个适合学习者特点的口译能力评测工具尤为重要。

针对中国英语学习者的特点，教育部和国家语言文字工作委员会于2018年研制开发了《中国英语能力等级量表》，其中专门针对口译能力开发了口译量表。量表是一种辅助评估的工具，通过描述任务的不同指标及不同水平上的详细表征，帮助评估者形成完整的评估意见(Stevens & Levi, 2013; 刘建达、韩宝成, 2018)。量表形态多样、功能多元，在各个学习场景发挥的作用不尽相同，学界对量表的功能也一直存在不同的见解。因此，在梳理量表的具体功能之前，有必要厘清几个基本概念，即量表是什么？包括哪些主要构件？本章将以口译初学者为研究对象，以本科阶段口译能力为起点，首先论述口译能力的构成、量表在口译能力评估中的作用，然后追溯量表的起源，分析量表的主要构件，聚焦量表的主要理论观点、量表助学的理论框架和基础，最后提出量表理论亟需回答的若干问题。

## 1.1 口译能力

### 1.1.1 口译能力的定义

口译能力的定义主要从性质、条件、结构三个视角出发，每个视角下的研究对口译能力的定义各不相同：

第一个视角为性质视角，即从心理学的角度入手，运用心理学中的能力原理推断口译能力的内在机制。20世纪90年代开始，口译学者（如 Hoffman, 1997; Kalina, 2000）依托认知心理学中的专业技能这一构念，提出口译能力和其他领域的专业技能一样，都包括如下特征：1）专业技能的发展经历了从对问题表面直观的理解到清晰的、概念上的、组织过的理解。2）在专业技能的发展过程中，学习者很少能够直接跳过某一发展阶段。3）随着练习的增加，该技能不再是学习者在有意的、不再费力的情况下获得的能力，这一技能不再是线性的，而是自动的模式识别能力。4）在教学过程中，熟练的从业者可以基于学生的技能水平预测其错误。由于性质视角下的口译能力定义依赖经验和直觉，评估权掌握在专家或熟练从业者手中，因此较少应用于学生自学和同伴评估。

第二个视角为条件视角，即从口译需要完成的任务或需要胜任的工作入手，将口译能力等同于胜任某一工作的条件的集合。这一定义的理论来源是职业教育中的能力本位教育理念（Competence-based Education），该理念认为能力的确定首先基于对目标市场的调查和分析，分析职业岗位所需能力的组成、学习和运用，并将职业岗位所需的技能和能力作为一切教育活动和职业资格评定的出发点和核心。

与性质视角下的口译能力相比，条件视角下的口译能力内容更加具体，层次更加丰富，评定指标也更加明确。要判断对方是否具备了某个级别的口译能力，只需让他在某一特定的职业条件下执行特定的口译任务，如能圆满完成，则可以评定他具备了该级别的口译能力。条件视角也决定了口译能力的评估方法，即组织真实的口译场景和任务，并让考生完成指定任务，由有经验的考官直接判断他们是否具备了胜任该岗位的条件。不过，条件视角与性质视角的评估一样，评估权集中在专家手中，学生由于缺少评估资源和评估经验，不能对自己或同伴的口译能力作出全面细致的判断（苏伟，2011）。

第三个视角为结构视角，即从口译能力的组成入手，强调能力构成要素的综合性。该视角认为口译能力是由多种元素复合而成的，因而也将口译能力称为综合口译能力。口译能力的结构视角沿袭自翻译

能力的结构观,学者通过建构多种翻译能力的结构模型为译员培训提供了理论基础。如 Bell (1991) 将翻译能力描述为包含多个要素的巨大综合体,包括目的语知识、文本类型知识、源语知识、主题知识、语言对比知识、交际能力。PACTE (2011) 认为翻译能力由双语能力、语言外能力、翻译知识、工具能力、策略能力五大子能力,以及心理生理要素构成,其中,心理生理要素还包括认知、态度、创造力、逻辑推理力、分析能力和综合能力等。

综合而言,目前,以结构视角为依据的学者对口译能力的综合性认识较为统一,但对其中的构成要素却存在着较大的分歧。在翻译研究领域,有学者指出,翻译能力的构件永远没有完整固定的答案(胡珍铭、王湘玲,2018)。译者需要根据时代的发展不断地扩展自己的翻译能力。在这个过程中,有些能力的重要性愈加凸显,而有些能力则随着科技的发展重要性逐渐降低。

口译能力的最新研究开始注重其阶段性特征,这是因为在口译学习的不同阶段,口译能力的培养目标和培养重点各有不同,尤其是对口译初学者而言,需要优先掌握口译能力的哪些要件在学界尚有争议,如鲍川运所言(2008:7),本科口译课程的教学是在一个不完整的条件下进行的,即学生是在语言习得尚未完成的情况下就开始学习翻译,学生的口译能力最高能达到什么程度,能胜任何种类型的口译,这是一系列很值得研究的课题。

### 1.1.2 本科阶段的口译能力

口译能力的发展路径有其阶段性特点,阶段性是指学习者口译能力的习得需要经历从新手级别到专家级别等不同阶段。口译能力的习得过程可以分为若干个阶段,每个阶段口译能力的特点各有不同,因此对应不同的培养目标和培养方式。Hoffman (1997: 199) 基于认知心理学的理论将口译能力的习得过程分为以下几个阶段:新手级(novice)、入门级(apprentice)、熟练级(competent)、专家级(expert)和大师级(master)。根据 Sawyer (2004: 72) 的观点,口译能力和口译课程的阶段性可以通过以下示意图表示:

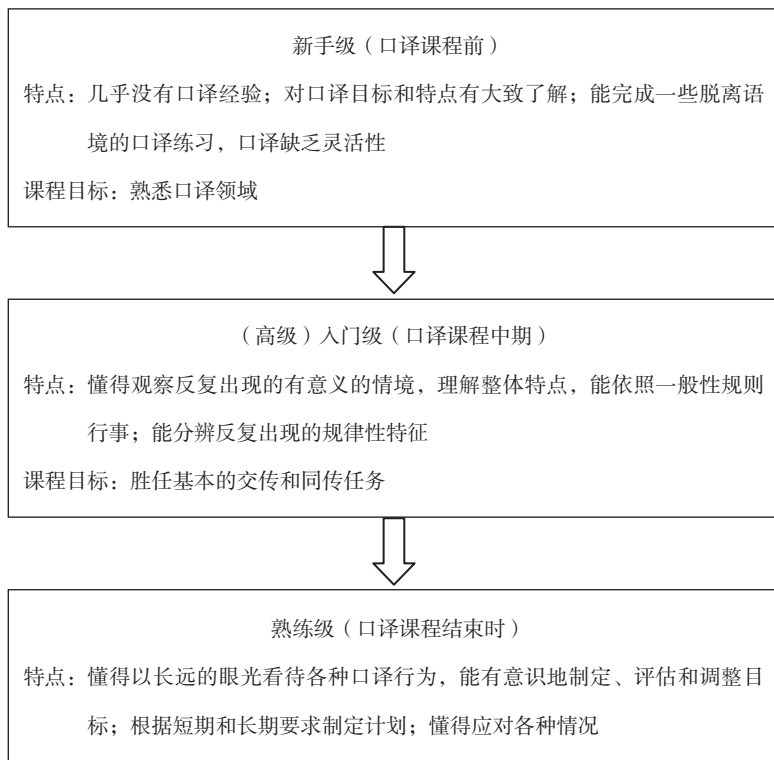


图 1.1 口译能力阶段性图解

Hoffman (1997: 199) 进一步解释说, 熟练级仅是学生在完整修完口译课程并顺利通过结业考核后所达到的能力级别, 只有再经过十年甚至更久的时间, 他们的口译能力才能达到专家级乃至大师级。需要说明的是, 上述口译课程主要是指西方国家口译项目研究生阶段的课程。

与西方国家相比, 中国的口译课程在发展历史、学习群体等方面有其独特性:

第一, 口译课程历史较短。在欧洲, 瑞士日内瓦 1941 年就成立了第一所高级翻译学院, 1965 年 12 月, 法国巴黎召开了首届会议口译教学会议, 探讨了口译培训的目的、方式、选拔测试、潜能测试、语言组合和语言水平等话题 (Mackintosh, 1999)。与之相比, 我国迟

至 1979 年才与联合国合作，在北京外国语大学设立了译员培训班，开设专门的口译课程，1996 年才由厦门大学发起和主办了首届全国口译教学研讨会。因此，我国在口译课程建设、口译能力培养等方面的理论和实践经验不够丰富。

第二，在学习群体方面，我国的口译学习者主要是汉语母语环境下以英语为外语的语言学习者，缺少欧洲国家口译学习者所拥有的天然的多语言环境和基础。鲍川运（2008）认为，我国本科阶段的翻译教学为我们提供了一个非常值得研究的课题。本科阶段翻译专业的课程应当进行的是翻译教学，目的是培养真正意义上的翻译人才，但是它又是在一个不完全的条件下进行教学，即学生在语言习得尚未完成的情况下就开始学习翻译，因此这给我们提出了一系列值得研究的课题，如翻译教学与普通外语教学相比，对语言习得能否产生更大、更积极的影响（鲍川运，2008：6-7）。

针对本科阶段口译能力的特征，《高校英语专业八级口试大纲》曾作出初步判断：本科阶段的口译能力包括中译英能力和英译中能力，其中中译英使用的语音材料是一篇中文演讲稿，语速约为 160-200 字/分钟，分段播放，翻译的原文字数约 200 字，可以做笔记。英译中的材料语速约为 100-150 词/分钟，分段播放，原文总字数约 150 词左右，可以做笔记。其中，口译能力的优秀级别界定为“能将讲话的绝大部分内容准确翻译出来，表达流畅”（高校英语专业八级口试大纲编写小组，2008：4）。可以看出，根据《高校英语专业八级口试大纲》（2008）的要求，此阶段的口译能力注重的仍然是语言的运用与转换，主要关注选词、谋篇、流畅三个层面，聚焦的是学生的双语能力和基本的语言转换策略。类似的，2020 年最新发布的《普通高等学校本科外国语言文学类专业教学指南》也提出英语类专业的口译课程目标是以英语语言能力为主，要求学生具备初步的口译素养并掌握口译的基本策略，因此“语言和内容基本准确、表达比较流畅”是大纲视角下本科阶段口译能力的培养目标（教育部高等学校外国语言文学类专业教学指导委员会，2020）。

本科阶段（也称基础阶段或初级阶段）口译能力的构成也得到了国内外学者的关注和重视：Schaffner（2000：154）较早论证了本科生的翻译能力，并将其定义为初级翻译能力（initial translation competence）。Schaffner 介绍说，以本科课程“翻译基本概念和方法”为例，该课程可以使学生了解翻译过程中进行系统分析和决策制定的理论基础，培养学生基本的翻译能力。因此，英国阿斯顿大学（Aston University）在本科一年级就给学生讲解翻译的语言学派、语篇翻译学和功能翻译学等基本概念，并在期末考试时专门设置面试环节，要求学生运用所学翻译理论简单地解释他们使用的翻译策略，这种课程设置取得的效果也非常明显。这一阶段的翻译能力表现为：在翻译时（主要是笔译）学生能主动考察语篇和上下文，思考选词造句的潜在目的、语篇的对象和体裁规定等方面（Schaffner，2000：149-154）。由此可见，经过培训后，学生能逐步从语言的微观层面（如选词）过渡到宏观层面（如语篇和体裁），培养初级翻译素养，本科阶段的翻译能力仍然以语言层面为主。

针对本科阶段如何达到语言和内部基本准确、表达比较流畅的口译能力目标，厦门大学口译团队于 2008 年提出了拓展版厦大口译训练模式，突出了两大模块，分别是口译准备（Foundation Building，简称 FB）和质量监控（Quality Control，简称 QC）。前者包括语言、知识和心理准备，后者包括测试和质量评估（陈菁、肖晓燕，2014）。口译准备包含两层含义：一是指长期的训前准备，即学员接受口译训练前需要奠定的语言、知识和技能等各方面的基础；二是指短期的以任务为中心的准备，即译员在接受口译训练后围绕任务进行的各项准备。关于语言准备，本科阶段学生译员的两个语言编码系统的转换要趋于自动化，能准确地理解原语并流畅地用译语表达。类似的，Wu & Liao（2018）提出的一次说一句（one chunk at a time）、Li（2013）总结的拖延（stalling）等策略也能帮助本科生尽快产出译语，实现表达流畅的目标。

### 1.1.3 口译能力评估

口译能力评估是口译研究中最具现实意义的内容，但也是研究者面临的最困难的问题之一。根据目的的不同，口译评估可分为：1) 职业能力评估，即判断受测者是否具备从业资格或具备获评某个职称、某种等级的资格，从而判断其能否胜任某个职业岗位或某项口译任务；2) 教学评估（或称训练质量评估），即判断学生是否具备某项能力，教学评估不仅要测量质量的实现程度，更要分析学生翻译能力的变化，其评估目的是要引导学生领会训练意图，从而达到训练目标（蔡小红，2007：138-139）。由于本书考察的是本科阶段的口译教学，而口译评估属于教学评估，因此本节将从评估指标和评估工具两方面来论述口译教学评估。

首先，评估指标方面，由于本科阶段口译能力的培养目标是语言和内容基本准确、表达比较流畅，因此内容、语言、表达应该是这一阶段的主要评估指标。Lee（2008）做了一项关于口译测试指标的调查。Lee邀请了9名职业译员做评分员，评估5名学生将英语口语译为韩语的口译成绩，学生的母语均为韩语，在此之前已经学习了一年的口译课程。评分员根据Lee拟定的评分指标评分，这些指标包括准确性、译语质量、表达三项，每一项分为0到6级别，其中0级为最低级别。三项指标的比重分别为40%、40%和20%。各指标的定义如下：1) 准确性，指如实传递原文内容，实现语意、语用的对等。2) 译语质量，指译语在语言上准确、自然、合乎语境。换句话说，它指的是译语的可理解性，即译语质量服务于听众，可再细分为语法、语音、词汇、句法、自然程度、语域和语体等方面。3) 表达，指公众演说或陈述技能，也就是有效的交际技能。评分员对译语表达的评估不必参考源语文稿。由于是通过录音记录学生口译的表现，评分员无法观察学生眼神交流、仪容体态等重要的公众演说技能，因此对表达一项的评分不包括眼神交流等内容。调查最后给出了评分员的反馈：大部分评分员觉得译语质量和表达这两个指标的界限比较模糊，不易分清，有50%的评分员觉得准确性这一项占的比重应为50%而非40%，即针对学生译员而言，内容应该是最重要的评估维度。类似地，Lee（2017）在最



新开发的专门针对本科生的口译评估指标中，也将内容（content）作为最重要的评估指标。关于语言指标，Riccardi（2002：121-125）特别强调，口译教学评估应当有别于职业口译评估，因为学生仍然处于学习阶段，不断提高语言技能仍然是他们的学习目标之一，因此教学中的评估考察的应该是学生是否完成了教学目标，哪些领域需要提高，哪一类错误最频繁等。

其次，针对本科生口译能力的评估工具，最常用的仍然是语料计量和评分量表两类。语料计量即对原文和译文语料进行一定程度的客观测量，依靠机器（如语法自动标注）和人工（如判断意义准确性）计算错误句子的数量，从而推测出译员的口译水平或能力。这种工具通过对每个错误句子扣分，或者对每个准确的句子（error-free T-units）加分从而得出最后的分数。该工具如同一台机械性的词汇测试机器，可以较大程度地保证评分的可靠性，但同时也把语言准确性狭义地界定在词汇对等、语法准确等层面，忽略了语篇语用等宏观层面，忽略了评分员的主观参与，因此效度存疑（Green，2014）。与之相比，评分量表给出了口译的整体特征（整体性量表）或各项目特征（分析性量表），强调了客观语料和主观判断的有机统一，尤其是经过量表培训之后，能够保证较好的效度和信度（Su，2020b），因此受到了越来越多的重视。针对我国本科阶段的口译学习者，教育部和国家语言文字工作委员会于2018年开发的《中国英语能力等级量表》口译子量表（五级和六级）就是目前国内针对本科阶段最新的口译能力评估工具。

## 1.2 量表的起源和发展

《现代汉语词典》将“量”定义为“用尺、容器或其他作为标准的东西来确定事物的长短、大小、多少或其他性质”（中国社会科学院语言研究所，1996：789）。量表，即以表格形式测量对象的性质，是一种具体的测量工具。按测量对象划分，量表也可以分为物理测量和心理测量两类，本书主要关注的是心理测量，即通过人的行为表现推测其特质和能力的间接测量。

在我国，心理测量量表起源于医学评估，古代医书《黄帝内经》中就描述了对太阴、太阳、少阴、少阳、阴阳和平五类指标的观察和评估，可看作测量指标的早期应用。量表正式大规模应用于考试测量是在科举考试中。以作文量表为例，考官对科场中所考的试帖诗划分了题意和格律声韵两个主要指标，以保证大规模选拔性考试中试帖诗的可比性和客观评卷的公正性（刘海峰，2010）。在科举制度发展后期，出题和评分指标日趋死板，出现了备受诟病的八股文标准。不过，也有学者认为，这一量表指标是汉字考试文体发展到高级阶段的产物，具有规范竞争引导备考，严定程式防止作弊，客观衡文快速评卷等优势。这些优势也正是量表统一性、公开性的体现（刘海峰，2010）。钱穆提出，科举最大的意义在于“用一个客观的考试标准，来不断地挑选社会上的优秀分子，使之参与国家的政治”（1996：405-406）。因此，量表在我国考试制度中的起源和发展集中体现了中央集权体制下公平选材、客观选人的社会现实。

英语中的量表（rubric）一词起源于13世纪的古法语 rubrique 一词，早期指书的章节标题，尤其是指示某一做法、流程、程序的提示语（霍恩比，2016）。随着19世纪西方心理测量学的发展，人们开始意识到人与人之间存在着心理上的差异。由于欧洲政治上民主思想的发展，教育与医学开始关注弱智儿童，智力测量也变得日趋重要。法国的比奈（Alfred Binet）与医生西蒙（Theodore Simon）于1905年合作编制的比奈—西蒙量表（the Binet-Simon Scale）是世界上第一个标准化的智力测量量表，该量表也成为西方心理测量中的重要工具。因此，量表在西方的应用与发展，在一部分程度上体现了西方社会尊重个体的民主思潮。

具体到语言力量表这一细分领域，美国政府部门——外事服务学院（Foreign Service Institute，简称FSI）于1955年发布的口语能力量表是较早开发的语言量表，考核的是驻外军事人员的口语能力。语言力量表发展的标志性产物是1990年代欧洲委员会（Council of Europe）主导开发的《欧洲语言共同参考框架》（The Common

European Framework of Reference for Languages, 又称《欧洲语言能力等级共同量表》), 该量表的开发初衷是为了顺应欧洲委员会语言政策部门对语言教学分级的要求, 希望把语言教学过程划分成若干个较小的、能独立授予学分的单元, 从而建立一个能得到欧洲各国相互承认和采用的共同参照指标, 使语言教学和测试透明化。

《欧洲语言共同参考框架》之所以成为语言量表发展史上的标志性产物, 其优越性主要体现在: 1) 在理念上, 该量表摒弃了传统量表将语言能力割裂为听、说、读、写四个孤立面的做法, 对语言能力进行整体和分类的描述, 考察使用者在接受、产出、交互、中介等语言活动中展示的语言能力; 2) 在研制过程上, 该量表以描述语为起点, 通过描述语采集、编写、建库、分级验证等方式, 为语言能力等级的描写提供了一个全面的框架, 在划分等级时力求量化和有序化; 这两个特点也影响了后续包括《中国英语能力等级量表》(2018) 在内的多个语言量表。

《欧洲语言共同参考框架》表于 2001 年 11 月由欧洲委员会通过并正式推荐给欧洲各国, 各国据此制定各自语言的教学大纲和开发教材, 并据此组织教学和语言测试。该量表推出后, 各地的语言测试认证机构纷纷与之挂钩: 欧洲语言测试学会于 2001 年决定将《欧洲语言共同参考框架》作为划分语言能力等级的指标。剑桥大学考试委员会推出的考试(包括商务英语考试、雅思考试、少儿英语考试、通用英语证书考试)所设立的水平等级都可以参照《欧洲语言共同参考框架》的等级指标进行解释。该量表在不同的测试之间架起了沟通的桥梁, 为不同的测试所报告的分数或成绩提供了比较的平台, 也为考试用户理解和使用不同的测试报告提供了方便, 有利于不同国家或地区相互承认各种语言能力证书(方绪军, 2007)。

在我国, 制定统一的语言力量表体现了我国的社会历史传统。我国的语言教育和社会发展总体特征一样, 也符合各地指标不一、发展不平衡的国情特征。各地缺乏统一的参照指标, 语言能力的定义或描述缺乏统一性, 具体表现在: 1) 语言能力描述参数设置不一, 比如

有些地区的教学大纲对“口头交际策略”这一表述作了描述，但其他地区却没有。2) 语言能力描述的精确度不一。对于能力参数（如语速、阅读速度等）各地的描述和刻度不尽相同。3) 语言能力（水平）的等级划分不一。语言能力等级少则两三个等级，多则十几个等级。综合而言，各地语言能力的描述具有经验性、相对性和主观性等特点（杨惠中、桂诗春，2007）。由此，制定统一的语言能力等级量表，可以在语言教学中提供一致的参照指标，使语言教学与测试的讨论基于共同的基础。

为了响应这一需求，在参考了《欧洲语言共同参考框架》等主流量表后，教育部和国家语言文字工作委员会主导研制了《中国英语能力等级量表》（2018）。该量表是首个面向中国学习者的英语能力的量表，经国家语委语言文字规范（标准）审定委员会审定通过，于2018年4月12日正式发布，作为国家语委语言文字规范自2018年6月1日正式实施。

与《欧洲语言共同参考框架》相比，《中国英语能力等级量表》（2018）在理念和研制过程两方面取得了新的进展：理念上，《中国英语能力等级量表》（2018）不仅涵盖了传统的接受和产出领域（如口头交际），还从语用能力和翻译能力（包括口译与笔译）的角度描述了英语能力；研制过程上，量表描述语在收集和编写时依据其作答情况补充标注了其质量层级，全面反映了不同描述语的功能和优势，最终收录的描述语优先来自质量层级中最高层级的描述语（刘建达、彭川，2017）。接下来，将详述该量表构件的特点。

### 1.3 量表构件

量表在英语中有多种相似表述，例如清单（checklist）、评分表（rating scale）、量表等。其中，清单是相对简单的一种表格形式，其构件为指标和选项，每个评分标准下一般只有“是”或“否”两个选项，评分人只需勾选其一即可，所以又被称为标准清单。清单可比做电灯开关，由于电灯只存在“开”或“关”两种状态，因此评估过程

便捷容易，评估结果简单直接。不少研究者（如 Green, 2014）指出，清单显示结果迅捷直观，便于理解，因此特别适合学习者的自评和互评，尤其适用于低水平学习者。不过，清单给出的“是”或“否”结果对语言特征的概括过于简化，最后学习者只能把所有二元结果简单相加得出最后结论（即有多少项符合，多少项目不符合），但不能给出更精细的信息。

与清单相比，评分表和量表的构件更加复杂，尤其是分析性评分表/量表，对各个指标给出了更加丰富的参数，以便更加精细地测量出不同受试的等级。这类表格可比作有旋转按钮的电灯开关，随着旋转圈数增加，电灯光线逐渐增强，即测量的结果不再只是“符合/不符合”的二元数据，而是更加精细更加完整的数据范围。

就评分表和量表而言，二者在很多情况下可以互用，都包括整体表（holistic form）和分析表（analytic form）两类，都注重对观察对象的全面测量。有时量表这一词较评分表更突出定性的描述与测量，即不是以具体的分数来描述某一指标下的受试表现，而是以具体细微的描述语凸显某一指标或等级的突出性特征。例如，评分表会选择以3、2、1等数字区分某一指标的不同表现，量表则以经常、有时、几乎不等描述语来区分不同表现（Stevens & Levi, 2013）。

量表的核心结构一般包括描述语、等级、典型活动三部分，以下将以《中国英语能力等级量表》（2018）为例，逐一介绍这三项内容。

### 1.3.1 描述语

描述语，即对学习者语言特征的陈述。描述语的来源包括专家的经验总结、对其他学科的描述语的借用、对实证数据的提炼三类。《中国英语能力等级量表》（2018）的描述语主要来自专家的经验总结、使用者（包括教师、学生、用人单位）的反馈、语言测试数据的提炼等渠道。该量表的描述语以交际能力的理论为基础，即视语言能力为学习者在一定的交际场景中得体地运用语言知识进行语言交际的能力（Bachman, 1990）。遵循这个理论框架，使用量表培养学生运用语言

进行交际的能力就是语言教学的中心任务，量表中描述语言能力的描述语也自然应该着眼于学习者能用语言完成怎样的交际任务，即所谓的“能做”描述（can-do statement）。这一描述语是对语言能力直接进行描述，可以直观地说明具有某种语言能力的学习者能用语言完成怎样的交际任务。从用户视角来看，对不同等级水平的语言能力进行直观的“能做”描述也便于用户理解和使用量表，理解量表所产生的成绩和结果。

量表使用“能做”描述固然有目标明确、使用便捷、促进学习等诸多优势，但“能做”句式本身并非没有局限：1）“能做”描述语本质上是对学习者目标水平的描绘，即英语能力目标水平符合哪个级别、哪条描述语的特征，正如设计者所说，量表不仅立足于中国学习者现有的英语水平，也着眼于学习者未来英语能力的发展，兼具描述性和规定性（刘建达、彭川，2017）。不过，“能做”描述语只能给出目标状态的特征，如果学习者水平不符合这一目标状态，量表本身并不能教会学生应该如何努力才能达到目标。例如，口译自评子量表第六级给出如下描述语：“我能监控译语的准确性及完整性，及时修正错误”（穆雷等，2020：146）。该描述语清晰实用，学生自评时能够迅速判断自己是否可以实现这一能做行为。不过，对于始终不得要领、不能实现该行为的学习者，量表除了给予其“不能做”这一单调的结论，并不能给出更多丰富的指导信息。2）“能做”描述语属于正面描述，给出的是学生语言各个层面的理想特征（expected traits）。学生在达到这一理想水平之前，由于学习习惯、学习风格和学习进度各不相同，总能遇到形形色色的学习困难，表现出各类负面的语言特征。仍以上述口译自评子量表第六级描述语：“我能监控译语的准确性及完整性，及时修正错误”（穆雷等，2020：146）为例，能够满足这一能做描述语的学生，表现出的是统一的、正面的口译特征。不过，仍然有一批无法实现这一描述语行为的学生，他们表现出的口译特征均不能达到这一目标，但各有缺点和不足：有的学生意识不到错误，有的学生无法区分明显和不明显这两类错误，有的学生无法在正确的时机及时纠正，

有的学生完全不懂得如何纠正。这些负面特质各自蕴藏了丰富的学习信息和诊断需求，但“能做”描述无法一一深入。

### 1.3.2 等级

《中国英语能力等级量表》(2018)的另一核心结构是等级。根据方绪军等(2008)的观点,语言能力等级量表的等级设计应该包括三步:1)设定量表的参照点,该参照点所代表的语言能力应该能够满足社会对学习者的语言能力的要求,参照点所表示的语言能力应该使用精细而清晰的描述(即量表的描述语),以便为确定其上和其下的语言能力等级提供有效和可靠的参照。2)设定量表的量度单位,确定语言能力等级的跨度大小及等级之间的分界。量度单位大小要合适,等级不宜太多也不宜太少,等级之间的跨度也应该等距。3)设定量表总长度,从实用、灵活、可操作的角度考虑,量表长度可以根据社会对语言能力描述和测量的需求情况,通过需求分析,确定需求量最大的语言学习者人群,从描述和测量这一人群的语言能力开始,向两端延伸,最终确定量表的总长度。《中国英语能力等级量表》(2018)最终将英语能力确定为九个级别,基本覆盖了基础教育到高等教育各个阶段的语言能力特征。

需要特别指出的是,据量表设计者介绍(王巍巍等,2018),口译教学对语言能力的要求较高,因此五级(包括高职、非英语专业大一、非英语专业大二、英语专业大一)以下的目标群体无法达到口译教学对纯语言能力的起点要求,口译量表能力等级划分以五级作为起点,包括五、六、七、八、九级这五个级别。

### 1.3.3 典型活动

量表除了给出描述语和等级外,还覆盖了各个典型场景或者语言活动。刘建达、韩宝成(2018)指出,典型活动即在某一特定领域围绕一定话题运用语言进行交际的典型行为,包括输入和输出文本信息等内容。典型活动包括话题和语境两个核心要素,前者是谈话或讨论的题目或主题是语言使用者通过各种口头或书面材料表达出的意义;后者指运用语言进行交际的情境,包括各种事件和说话人的内外在情

况等各项要素。量表设计者评估了语言学习者最常接触、最实用的典型活动，并根据话题功能提出了六类典型活动，分别为叙述性、描述性、说明性、论述性、指示性、交流性活动。每类活动的筛选和排列遵循以下规律。

首先，典型活动并不是简单的罗列，它们按照一定的顺序，遵循语言能力习得顺序依次排列。以记叙性活动为例，在听力子量表中，初级学习者（第一级）只能听懂绘本故事、动画片等简短的听力文本。随着听力能力不断提高，能够听懂体育赛事播报、现场新闻报道、电视访谈等难度较大的内容。

其次，典型活动的难度设计主要集中于几个有限的特征性参数。参数之一是内容的抽象程度，不同阶段的学习者所能理解和表达的话题内容会随着认知能力和社交能力的增长而发展。他们所能理解和表达的内容慢慢从自我世界、眼前世界延伸到更广阔的世界，话题内容变得越来越抽象和复杂（刘建达、韩宝成，2018）。参数之二是语言特征。较容易的活动文本结构清晰、用词简单，但难度更大的活动往往文本结构复杂、术语生僻词多、观点隐晦、专业性强，会含有隐喻和双关等修辞和概念抽象的材料（何莲珍、陈大建，2017）。

### 1.3.4 小结

本节主要介绍了量表这一评估工具相对其他评估工具的结构特点，并以《中国英语能力等级量表》（2018）为例，述评了量表描述语、量表等级、量表活动三大主要内容：量表描述语即语言使用的特征陈述，以“能做”描述为主要表现形式；量表等级为一至九级，共九个级别，覆盖了从基础教育到高等教育各个阶段的学习者，能够比较精细而全面地刻画语言能力发展的阶段性特征；量表活动分为叙述性、描述性、说明性、论述性、指示性、交流性活动等类型，按照活动涉及话题的熟悉或抽象程度、话题文本的语言特征等参数设置不同的难度等级。

需要指出的是，获得量表并不意味着学习者能够自主开展学习和评价活动，量表的助学助评功能仍受到多种条件的制约，正如



Panadero & Romero (2014) 所述, 量表与学习的关系动态多元, 即使配备了量表, 学习者也会因压力过大而采取规避策略, 从而制约量表的助学效果。就《中国英语能力等级量表》(2018) 而言, 经过严谨规范的开发研制, 其助学助评效果在实际操作中能否顺利实现, 会受到哪些因素的影响, 这也是本书着重探讨的内容。

## 1.4 量表的理论基础

量表的助学功能体现在调节学习者的情感和认知, 调节学习者的社交功能等方面, 其理论基础主要包括自主学习理论和社会文化理论两类。

### 1.4.1 自主学习理论

量表一个重要的理论基础就是自主学习理论。早期的自主学习理论关注点在课外的自学活动, 如 Holec (1981) 从自主学习的能力出发, 将自主学习视为学习者有能力制定学习计划、选择学习材料、监控学习过程、自我评价学习结果等方面, 强调课外自主学习的重要性。而 Dickenson (1987) 从学习的情境出发, 将自主学习定义为学习者要对与其学习有关的所有决定和这些决定的后果承担全部责任的学习情境。这两类理论视角都突出了学习者自己主导学习的特征, 但没有对如何实施、如何考核, 尤其是量表如何在自学中发挥作用进行系统论述。

随着认知心理科学的发展, 学界对语言学习本质的认识不断深入, 研究者开始探讨自主学习过程中的心理和情感状态, 尤其开始关注量表对自主学习的推动和调节作用。

首先, 量表是推动自主学习的一个重要途径。学习者的自主学习或自主评估主要有三个途径: 直接评测、使用量表、使用指导手册 (Panadero & Romero, 2014)。其中, 使用量表成为促进自主学习的一个重要途径, 这不仅仅是因为量表规定了各个表现指标及其内涵, 明确了学习者的努力方向, 还因为量表大都给出了具体的评估策略, 学习者可以通过实践这一评估过程, 更清楚地了解自身的优势和不足,

确定自己的努力途径。量表的推动作用还表现在，它能为学习者的成绩提供诊断性反馈，这也是量表最直接、最常见的助学功能。例如，参照《欧洲语言共同参考框架》开发的一项基于计算机的诊断性语言测试（Diagnostic Language Tests，简称 DIALANG）就充分体现了量表对学生语言能力的诊断反馈功能。学生完成 DIALANG 测试后，项目会对考生的自我评估等级和 DIALANG 测试成绩等级进行比较分析，给出自评和机评的差异，鼓励学生去探寻为什么会存在这样的差异。这些反馈信息能够提高学习者对语言本质和语言学习的认识，从而帮助他们更准确地诊断自己的语言能力。此外，项目还对学习者如何从目前的能力等级向更高一个等级发展提供了具体建议，以鼓励学习者去思考该采取什么样的行动来提高自己的语言能力，这些反馈均能够帮助学习者认识自己的优缺点，促使学习者比较自己对语言能力的期望与量表等级之间的差距，激励他们采取相应的措施提高语言能力（Alderson, 2005）。

其次，量表是调节学生情感和认知的重要手段。心理视角下的自主学习研究认为，学习者需要不断调节自身的情感和认知状态，提升自我效能，促进学习效果，而量表就是情感或认知调整过程中一个重要的干预变量。自主学习能力意味着学习者能独立做出并实施影响自己行动的选择能力和意愿，而这种意志取决于学习者的动机和信心程度。陶伟（2017）通过心理测量问卷发现，语言自主学习行为与学习者的倦怠程度相关系数较高。Reynolds-Keefe（2010）发现，使用量表能极大地减轻学习者焦虑，提升信心，平衡精力分配，从而提高学习效率。

量表不仅仅能调节学习者的情感，还能调节其元认知能力，这里的元认知能力主要包括自我规划和修正、自我监控和反思的能力。由于量表的指标清晰、操作性强，学生可以根据其要求事先规划自己的学习策略，也可以根据评估后的结果进行事后反思并调整学习计划。汪琼等（2019）发现，课堂上使用的量表可以成为学生的一种反思支架，帮助学习者结构化地运用课程所学知识对自己或他人的作业作出

评价，因此，运用量表进行作业评价的过程，就是审辨和反思的过程。特别需要提出的是，量表评估不仅能促进被评估者的审辨和反思，还能促进评估者的思考。汪琼等（2019）还特别指出，互评活动是一个健康的学习生态，参与互评的学习者在这个过程中都会认真参与，评估者写下的不同长度的评语表明评估者也有进行反思活动，只是反思意识的意愿和强弱可能因人而异。

关于量表促进元认知能力的具体测量方法，汪琼等（2019）首先将反思意识操作化定义为“学习者在进行作业互评的时候，有无应用作业互评量表的意识”。由此对量表促学效果的判断可以立足于学习者的互评记录。该研究进一步提出了评语与量表原文在字符上的匹配程度这一视角，即学习者在写评语的时候如果能够借鉴量表的语言，如果评语文本与量表高度匹配（使用量表原词或者近义词），就代表学习者在写评语的时候在思考作业的某个方面应该达到的水准，能表现出一定的反思意识，且这样写出的评语会有一些的针对性。对被评者来说，这样的评语也会有参考价值，因此会被界定为高质量的评语。依循这一研究视角，该研究发现，25.5%的学习者其评语和量表之间没有吻合之处，一定程度上说明了这些学习者在写评语时没有体现出反思意识。不过，成绩好的学生互评的时候会更多参考量表，生成的评语质量较高，而且他们作业得到的评语也质量较高，因为其他同学愿意分析和借鉴这样的作业，其评语也会大量引述和论证量表的各个维度。

具体到《中国英语能力等级量表》（2018），根据负责人介绍，研制《中国英语能力等级量表》的目标是促进中国学习者自主学习，这既包括设置学习目标，借助量表提供的指标参照，确定自己的学习计划，也包括学习者自我监督学习过程，依据量表选择学习方法和策略，利用形成性评价手段来管理自己的学习过程（刘建达，2017）。由此，量表对学习者的自我学习发挥着积极作用，通过描述语支持学习者进行自我评估，提供以学习者为中心的语言学习方法，鼓励自主学习（Runnels, 2016）。不过，量表的各个结构对自主学习的具体促进作用和实现方式仍然需要进一步的研究。

### 1.4.2 社会文化理论

量表助学的另一个重要理论基础是社会文化理论 (Vygotsky, 1980)。社会文化理论提出, 个体知识的获取和发展有赖于个体在社会活动中的交流。在交流中, 学习者在与周围人的互动和合作中激活了自身的内部发展程序。由于交流双方在知识、经验、能力上不尽相同, 交流成为一个搭建脚手架 (scaffolding) 的过程, 能力较弱者在能力较强者的帮助下能够完成自己无法独立完成的任务。学习者现有的能力水平与在他人帮助下可以达到的水平之间的距离被称为最近发展区 (Zone of Proximal Development, 简称 ZPD)。当学习者能够独立完成在他人帮助下才能完成的任务时, 其最近发展区会关闭, 进入下一个最近发展区。学习可视为从一个最近发展区向下一个最近发展区发展的过程。

量表作为一种特殊的脚手架, 一方面为学习者交流提供了对话的脚本, 促进了不同类型学习者之间的有序沟通, 另一方面也为不同水平学习者之间的同伴反馈提供了一个公平竞技场 (level playing field): 质量指标不再是高水平学生的独有话语内容, 低水平学习者也能参与讨论、发表意见, 也能就同一质量指标在不同水平学生上的体现提出自己的见解, 向对方的学习习惯、学习风格提出基于量表的建议和意见。由是观之, 量表不仅提升了低水平学习者的信心和参与度, 更重要的是组建了一个受众更多、参与面更广、交流更活跃的学习共同体, 真正发挥了帮助不同类型、不同水平学习者的脚手架作用。

不过, 量表在互评过程中的使用也面临挑战。周季鸣、束定芳 (2019) 发现, 学生使用量表互评的主要困难包括理解困难、使用困难、认知困难三个方面。量表理解困难方面, 有学生提出对某些宏观指标, 如语域 (register) 感到困惑, 最终没有依据这一指标给同学的英语作文评分, 也未得到与教师进一步沟通、了解语域这一内涵的机会。认可困难方面, 有同学提出某些指标的设置, 如流利 (fluency) 极不合理, 因此互评时并不使用这一指标。最后, 即使理解并充分接受了量表指标, 学生仍反馈评价时要在短时间内考虑语言、内容和报

告方式等维度，在认知上存在巨大压力，因此无法顺畅地使用量表完成互评。针对上述困难，除了加强量表指标解释、对量表内涵的沟通外，在量表使用过程中教师应给予方法指引，如递进式评估（一次只关注一项指标，逐项累积）、匿名评估、优缺点同步评估等。同样，徐鹰、章雅青（2020）在学术英语课堂推行基于量表的同伴评估中也发现，同学集中于表面的语言特征，比如作文的字迹是否清楚、用词是否丰富、句式是否复杂、语法是否正确等，而且多是褒奖的评语，缺乏对于思想的逻辑性、论点的说服力以及学术写作规范性等方面的建设性意见，这可能直接导致其反馈得不到同伴的信任，从而使得量表的社交效用大打折扣，因此迫切需要教师进行干预或量表培训，以挖掘量表的助学潜力。

## 1.5 小结

本章介绍了本科阶段口译能力的评估目标，梳理了量表的起源和发展，并以《中国英语能力等级量表》（2018）为例，评述了量表的典型构件，最后着重述评了量表助学的理论基础，即自主学习理论和社会文化理论。前者侧重量表对个人认知和情感的调节，后者侧重量表对同伴社交的指导和支持。本章发现，总体而言量表在自评（自主学习）和互评（同伴社交）领域均得到了较多的正面证据，但研究者也提出，仅凭量表本身并不能保证其助学效果能够充分发挥，量表需要和其他干预方式（如量表培训）相结合才能充分发挥效用。本书将针对《中国英语能力等级量表》（2018）这一新量表，以国内本科阶段的口译学习者为研究对象，考察其在不同培训方案下的助学效果。此外，现有研究多关注自评和互评这两类场景，探究不同背景的学习者使用量表的差异，本书将以英汉口译这一跨语言交际任务为窗口，引入不同背景的教师（即中文母语教师和英语母语教师），综合考察以学生为主体，多背景教师为辅助的量表应用场景，从而提出更加丰富全面的量表应用方案。