目　录

表　目

图　目

第一章
导论

1.1 研究背景

1.1.1《中国英语能力等级量表》

语言能力量表相关的研究已经有 60 多年的历史，主要集中在北美、欧洲、澳大利亚等国家和地区。其中，《欧洲语言共同参考框架：学习、教学、评估》（*Common European Framework of Reference for Languages: Learning, Teaching, Assessment*）（简称《欧框》）是很有代表性和影响力的语言能力量表，它为语言能力的描述提供一个等级标准框架，同时使用者可以根据自身需要进一步细分等级，具有一定的灵活性（邹申等，2015）。《欧框》的研制吸收了当代语言学、语言教学和语言习得等领域的研究成果。自《欧框》问世以来，我国学者便一直关注和研究《欧框》，相关研究包括对《欧框》的解读和介绍（韩宝成，2006；张建琴、邹为诚，2010；付桂芳、Broeder，2011）以及分析、对比和应用研究（方绪军，2007；白乐桑、张丽，2008；岑海兵、邹为诚，2011；黄婷、贾国栋，2012；刘静观，2012）。同时，我国学者开始呼吁建立中国的语言能力标准，并提出诸多的建设性建议（杨惠中、桂诗春，2007；杨惠中等，2012；Jin & Wu，2014）。

2014 年 9 月，国务院印发了《关于深化考试招生制度改革的实施意见》（下称《意见》），明确提出要加强我国外语能力测评体系建设，

标志着新一轮考试招生制度改革全面启动。这是我国教育领域一项具有基础性、全局性意义的重大制度改革。为了贯彻实施《意见》，教育部于2014年10月启动“国家外语能力测评体系建设”项目。作为该项目的一项基础性工作，《中国英语能力等级量表》（下称《量表》）的研发也同时启动。目前，《量表》已由教育部和国家语言文字工作委员会发布，并于2018年6月1日正式实施（教育部考试中心，2018）。截至目前，已有不少研究从不同的方面介绍了《量表》建设的过程和成果（刘建达，2015，2019a；刘建达、吴莎，2019；金艳、揭薇，2017；朱正才，2015，2016；邹申等，2015）。

《量表》由六部分组成，分别描述语言理解能力、语言表达能力、语用能力、中介能力、语言知识和语言策略。语言理解能力包括理解口头语言信息的能力（听力能力）和理解书面语言信息的能力（阅读能力）；语言表达能力包括用口头语言表达信息的能力（口语能力）和用书面语言表达信息的能力（写作能力）；语用能力分为语用理解能力和语用表达能力；此外，《量表》还描述了中介能力，即口译和笔译能力。各类语言能力都是围绕描述、叙述、说明、论述、指示、互动六项功能所需的语言能力进行描述。《量表》的语言知识框架包括语言组构知识（即语法知识、篇章知识）和语用知识。《量表》的语言策略指的是语言使用者完成一项语言交际活动所采取的有组织、有计划、目标明确的行动步骤，具体包括规划、执行、评估、补救四个步骤（刘建达、韩宝成，2018；刘建达、吴莎，2019）。

1.1.2《中国英语能力等级量表》之口语量表

《量表》之口语能力量表（下称口语量表）从口语能力、口头交际策略和口语文本特征等多维度描述各等级英语口语能力的典型特征。其研发的总体路径是基于中国英语学习者口语交际能力理论模型，在交互理论的指导下界定口语能力构念，采用文献法、采样法以及专家论证等方法撰写和改编描述语，建立口语能力描述语库并划分等级，采集全国大样本教师和学生评价数据为等级划分获取依据，提出具有区别性特征的口语交际能力等级描述。

已发布的口语量表由两部分组成，分别描述口头表达能力和口头表达策略，包括口头表达能力总表、口头表达策略、口头表达能力自我评价量表等 12 张量表，分为基础、提高、熟练三个阶段，设立九个能力等级，总共 379 条描述语。

口语量表将为我国英语口语的教学、测评和学习提供一个统一的参考标准。口语量表对教学的指导作用体现在进一步明确英语口语教学的目标，有利于教学的组织和衔接，有助于教师把教学划分为较小的、具有明确目标的阶段，为各阶段选择合适的教学方法，更有效地组织课堂教学。同时，量表也将有利于教材的编写和开发，提高教育系统的透明度。

口语量表对测评的指导作用体现在为英语口语考试的研发提供科学、系统、清晰的评价指标体系，有助于教师设计检测学习成果的评估办法和工具，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。口语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。考试机构可以通过考试与量表的衔接实现不同考试之间的可比性，增强考试的透明度，以便于考试用户更好地理解和使用考试结果，有利于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，还将有助于我国的英语口语测评与国际标准接轨。

口语量表对学习的指导作用体现在帮助学习者了解自己的知识掌握程度和技能达到的水平，跟踪自己的学习过程，开展自我评估，从而进一步明确学习目标，思考和调整自己的学习策略。此外，口语量表有助于学习者根据学习目标，挑选合适的学习材料，确定适合自己的学习内容、任务和测评，提高学习效率。

1.1.3 口语量表的效度研究

量表的用途广泛，因而其效度尤为重要。语言能力量表只有经过效度论证，才能证明其有效性和可操作性（杨惠中等，2012）。效度是评价量表质量的重要标志。因此，口语量表的效度研究贯穿研发的始终，且在口语量表发布后，依然需要不断搜集证据，证明口语量表在

实际应用中的效度。

纵观教育测量和语言测试领域，效度理论的发展脉络大致经历了分类论、整体论和论证论三个阶段（Chapelle，1999；韩宝成、罗凯洲，2013）。效度理论的发展推动和影响了考试和量表的开发和效度验证模式，特别是效度论证论，研究者将自己的研究结果解释为在论证的基础上支持特定假设，有助于研究者明晰研究方法，为不同的研究主题提供共同的以论证为基础的效度研究元素。效度论证论除了广泛地用于考试的效度验证研究中，还被用于评分量表的开发和验证研究，如 Knoch（2007）、Knoch & Chapelle（2017）提出的基于论证的表现性测试评分量表的效度验证框架。

能力量表的效度有别于评分量表以及考试的效度，三者本身是不同的研究对象。因此，口语量表的效度验证有别于口语考试和评分量表的效度验证。为保证口语量表的稳定性、有效性和生命力，其效度验证需要建立一个科学的、完整的、有针对性的验证框架，不能完全沿用考试的效度验证框架。

目前，国内外关于能力量表效度验证的研究较少，且缺乏系统连贯的理论框架指导，即整合各个方面的经验证据或理论依据的能力量表效度论证框架。存在这样的现状有诸多原因：一方面，能力量表的效度研究不如测试的效度研究那样悠久，世界各地对于能力量表的关注度也是近年才逐渐升温；另一方面，能力量表开发者比较注重量表的研制和开发效度，对能力量表效度的检验往往更关注内容效度和结构效度，很少有研究将量表的效度延伸或拓展到量表的外部效度和后果效度。量表的外部效度是指量表的外延性，量表的后果效度检验指通过运用量表进行效度检验，即量表开发基本完成以后进行的后验。此外，量表的开发者和量表的使用者往往从不同的立场和角度看待量表，从量表的开发到使用，效度验证存在断层，量表的开发者不一定是量表的使用者。如 ACTFL 和《欧框》之类影响深远的语言能力量表，在其开发伊始和投入使用后的很长一段时间，学界对于其质量和应用研究的关注度一直很高，推崇者有之，批判者有之。即便如此，

目前还没有研究系统地开展这些量表的效度检验。

能力量表的效度验证框架需要综合各项指标，以整合取自各个方面的经验证据或理论依据（朱正才，2016）。因此，口语量表的效度验证是复杂和多维的，需要在效度验证框架的指导下，长期地、动态地开展，不断收集和积累各个方面的效度证据，通过具体的主张和理据把效度研究框架中各个重要属性有机地联系起来，并明确各个主张和理据的证据来源，从而有助于口语量表核心成分的解释，并探讨口语量表使用的后效。

1.1.4 口语量表的效度取证

《教育与心理测量标准》（*Standards for Educational and Psychological Testing*）（简称《标准》）指出，合理的考试效度论证是要将各种证据整合成连贯的体系，论证的可靠性取决于证据和理论在多大程度上支持考试成绩的预期释义及用途（AERA，APA & NCME，2014：21），可以说效度研究就是收集尽可能多的证据去支持考试（谢小庆，2013）。虽然《标准》源自美国，但其影响力已经远超北美各国（Zumbo，2014）。在开展语言测试的效度验证时，研究者需要依据效度论证框架采集证据和报告研究结果，如果这一技术问题没有解决，那么效度验证最后可能就会流于核对清单式（checklist）验证（Fulcher，2015：116-120），而不是真正的效度论证（Chapelle *et al.*，2010）。

口语量表的效度需要来自理论和实践各个角度的证据支撑：第一、来自口语量表内部的证据，如口语量表的构念和级别效度；第二、口语量表的外部效度证据，如量表与考试之间关系的证据；第三、口语量表的后果效度证据，口语量表的使用和影响，如在考试或教学中的应用。口语量表开发过程中，内部效度证据是非常重要的。North & Schneider（1998）、De Jong（1990）认为，基于专家的经验和直觉或者借鉴现有的量表而开发的语言能力量表在实践中会产生各种问题，特别是在高风险的环境中使用这样的能力量表。Fulcher（1996）回顾了 ACTFL 量表效度的研究，指出如果量表在开发伊始缺乏扎实的实证基础，就很难在其投入使用后对其进行效度研究，研究结果可能

与量表的假设及构念完全背离。量表初步构建后，研究者需要采集基于量表内部结构的证据，可以使用统计方法如因子分析或者结构方程建模分析量表的结构，检验量表等级描述语的可量测性（scalability）（Henning，1992）。量表发布后，研究者要进行科学的研究设计和数据分析，收集量表的外部效度和后果效度证据，为使用和改进量表提供依据。

总之，口语量表的效度取证是效度验证的关键，必须有效地收集和正确地使用口语量表效度证据。同时，国外有关能力量表的相关研究也可以为口语量表效度研究提供参考。

1.2 研究目的和内容

本研究是口语量表研制的一个组成部分，目的是构建口语量表的效度验证框架，并应用该框架综合分析口语量表的效度。本研究的内容包括以下三个方面：

第一，理论构建。本研究旨在从效度理论的最新视角构建量表的效度验证框架并将该框架应用于口语量表的效度研究，探索口语量表效度验证的内涵，发展和丰富语言能力量表的效度理论研究。

第二，实证研究。本研究将根据所构建的效度验证框架，收集相应证据，分析各方面的证据，从而将多个效度研究视角有机地整合到口语量表的效度论证中。此外，本研究尝试使用不同的研究方法，使不同效度证据的相互关系通过实证检验得到解释。

第三，应用探索。本研究将探索如何在实际应用中进一步论证量表的效度，从而将口语量表的效度研究范畴拓展到后期的实际应用阶段。

在研究的总体设计上，本研究采用了定量、定性和经验方法，既基于较为成熟的语言测试效度理论并吸收语言测试效度验证的长处，同时结合了本研究的需求和现状，利用现代测量学技术对口语量表进行系统、深入的效度检验。