

# 前 言

英语教师在日常教学工作中需要不断依据测评结果做出相应决策。以1名英语教师执教50名学生为例：一次测评结果将生成50个决策，而一年中大大小小的各类测评极有可能需要这名教师做出上万次的决策，因此，教师的工作离不开测试与评价。具备测评素养是做出正确、有效的决策的前提，这已经成为了教师应具备的基本知识和重要技能。国际上，测评素养的相关研究已经从普通教育学领域逐渐发展到了教师资格认证、教师发展、语言测试等诸多方面。在我国，对测评素养的研究和关注也处于日益高涨的上升期，有关学者结合国际上的研究成果正逐步探索符合我国国情的途径和模式。

本书为国家社科基金项目“中国高校英语教师语言测评素养研究”（16CYY028）的最终成果之一。本书围绕“语言测评素养是什么”“语言测评素养应如何构建”“我国高校英语教师语言测评素养的现状如何”“语言测评素养应如何衡量”“语言测评素养应如何提升”等问题展开。全书共九章。

第一章为“语言测评素养的基本理论”。这一章主要围绕语言测评素养的概念与基本理论展开。语言测评素养的研究起源于教育学领域，并在语言测试界中得到了长足的发展。该章首先从语言测评素养的定义出发，梳理这一概念的本质与内涵；接着逐一列举并解释语言测评素养研究中的四个模型，并对这些模型加以评述。

第二章为“教师测评素养国际研究热点与趋势分析”。这一章采用可视化图谱分析法，对国外教师测评素养进行文献计量学分析和梳理。该章发现国外教师测评素养研究处于发展期，研究热点主要集中在课堂测评、学生学习、教师发展等方面；研究视角主要围绕教师测评素养需求调查、水平调查、利益相关群体界定和内涵构建等不同维度开展。该章指出，教师测评素养是教师发展和语言测试相交融的研究重点，也是我国未来外语教育深化改革和教师核心素养发展

的关注要点。

第三章为“语言测评素养：语言测试专家的视角”。这一章认为，随着语言测试研究的深入，语言测评素养的构念也进入新的发展阶段，沿用原有的语言测评素养构念可能无法适应如今英语教师发展的需要。本章是在第一章和第二章基础上开展的一项实证研究。通过深度访谈语言测试领域的专家学者，该章提取出语言测评素养构念发展的新动态和新外延，梳理出我国培养语言测评素养的盲区等薄弱环节，并提出了语言测评素养构念的修正模型。

第四章为“语言测试课程研究”。作为培养并提升语言测评素养的重要渠道，语言测试课程的地位可见一斑。特别是在本科生和硕士研究生的培养过程中，语言测试课程已经逐渐成为师范类英语专业学生以及英语相关专业硕士生的重要课程。该章从语言测试课程的教学大纲入手，梳理并分析了语言测试课程在我国语言测评素养中的基本状况，并结合第一章中的有关内容指出现有语言测试课程的不足之处，提出了今后对其改进的一些建议。

第五章为“语言测评素养量表：回顾、比较与展望”。这一章从语言测评素养本身转向测量这些素养的标准或量表。该章指出，教师测评素养作为核心主题不断出现在教师发展和语言测试与评价领域的研究中，从术语演变、需求调查、水平调查、利益相关群体界定和内涵构建等维度，呈现了跨学科、跨领域、多视角、多维度的教师测评素养研究范式，量表的开发和使用是其中之一。该章以美国、英国、澳大利亚、新西兰为代表的英语国家和欧洲地区教师相关专业标准文件和指南为主要研究内容，对比分析了它们的异同，对笔者所研制的《中国高校英语教师测评素养量表》有很大的参考价值。

第六章为“英语核心素养指导下的语言测评素养”。该章是一项基于师范类院校毕业生在从事小学阶段英语教学活动中所反映出的测评素养研究。该章基于我国英语学科核心素养的构成及其内涵，从师范类院校毕业生的测评素养入手，梳理并分析了我国部分地区小学阶段英语学科终结性测试的形式与内容，并着重对英语学科核心素养的内涵与终结性测试的指向进行关联研究。本章认为，我国师范类院校毕业的小学英语教师在英语学科终结性测试中深受结构主义的影响，极少测量英语学科核心素养的真正内涵与具体表征，一定程度上导致了学习英语就等同于学习英语组构知识的误解，降低了学习者的学习兴趣与动机。

第七章为“语言测评素养与《中国英语能力等级量表》”。该章在回顾《中国

英语能力等级量表》(CSE)的各项重要元素后指出,要真正落实量表的相关规定,语言测试领域的研究人员、教育管理者、一线教师等都应对此深入学习并统一认识,提升自我测评素养。该章在详细介绍 CSE 写作分量表研制细节和框架结构的基础上,指出可从 CSE 中引申出语言测评素养的基本要素。该章最后指出,在 CSE 的实施背景下,语言测试的利益相关方需要掌握哪些测评素养内容和原则。

第八章为“构建《高校英语教师语言测评素养量表》”。这一章在前几章的基础上,结合我国高校英语教育的实际情况,推出了《高校英语教师语言测评素养量表》。该量表不仅包含不同群体的高校英语教师,也基本涵盖了语言测评素养各个维度的内容,是本书前几章文献分析和实证研究的综合成果。高校英语教师可通过自我测评的方式得到其测评素养的概貌。此外,该章也介绍了这一量表效度验证的结果,说明其构念效度、可操作性等问题。

第九章为“语言测评素养的常态化培养与提升”。这一章主要阐述了语言测评素养量表的应用。出于语言测评素养常态化培养和提升的考量,本书提出将量表嵌入到手机移动端 APP 之中,并通过高校教师的自评结果来推送一些他们自评后感到薄弱的环节。APP 的设计包含量表自评、知识版块、互动讨论等专区,为语言测评素养的提升提供了长效机制。基于 APP 所形成的大数据,本章也提出了今后进一步进行语言测试素养提升的研究课题等。

本书的出版受到上海市 I 类高峰学科(外国语言文学)建设项目以及上海外国语大学青年教师科研创新团队“人工智能发展中的语言习得与语言测评前沿研究”(项目号:2020114050)的资助。在本书撰写过程中,笔者得到了多位前辈学者和同行专家的指点。笔者衷心感谢上海外国语大学邹申教授、上海交通大学金艳教授、对外经济贸易大学江进林教授等专家的帮助和指点。此外,英国贝德福特大学(University of Bedfordshire)的丽兹·汉普-莱昂斯(Liz Hamp-Lyons)教授也在文献收集、数据解读等方面提供过莫大帮助。在本书付梓出版之际,作者也向上海交通大学出版社的臧燕阳编辑以及其他编辑同事表示由衷的感谢。

# 目 录

<b>第一章</b>	<b>语言测评素养的基本理论</b> ·····	<b>001</b>
	第一节 语言测评素养的定义·····	001
	第二节 语言测评素养的发展·····	003
	第三节 语言测评素养的外延·····	005
	第四节 语言测评素养的模型·····	006
<b>第二章</b>	<b>教师测评素养国际研究热点与趋势分析</b> ·····	<b>015</b>
	第一节 教师测评素养简述·····	015
	第二节 基于文献计量学的测评素养研究设计·····	016
	第三节 计量结果和数据分析·····	018
	第四节 对我国外语教师测评素养发展的启示·····	031
<b>第三章</b>	<b>语言测评素养：语言测试专家的视角</b> ·····	<b>035</b>
	第一节 语言测评素养文献再探·····	035
	第二节 专家视角的研究设计·····	038
	第三节 专家视角的研究发现与讨论·····	040
<b>第四章</b>	<b>语言测试课程研究</b> ·····	<b>045</b>
	第一节 语言测试课程研究简述·····	045
	第二节 语言测试课程教学大纲的比较研究·····	048
	第三节 语言测试课程的短板和发展·····	055

<b>第五章</b>	<b>语言测评素养量表: 回顾、比较与展望</b> .....	<b>059</b>
第一节	教师测评素养历史回顾 .....	059
第二节	各国教师教育和测评标准类文件的比较研究 .....	061
第三节	语言测评素养量表的展望 .....	080
<b>第六章</b>	<b>英语核心素养指导下的语言测评素养</b> .....	<b>085</b>
第一节	英语学科核心素养与青少年英语测试 .....	085
第二节	英语师范生在命题实践中的测评素养 .....	088
第三节	对师范类毕业生测评素养培养的启示 .....	092
<b>第七章</b>	<b>语言测评素养与《中国英语能力等级量表》</b> .....	<b>095</b>
第一节	《中国英语能力等级量表》: 以写作能力为例 .....	095
第二节	《中国英语能力等级量表》中语言测评素养的基本要素 .....	105
第三节	《中国英语能力等级量表》指导下的语言测评素养 .....	112
<b>第八章</b>	<b>构建《高校英语教师语言测评素养量表》</b> .....	<b>115</b>
第一节	构建测评素养量表的原则与考量 .....	115
第二节	《高校英语教师语言测评素养量表》的开发与验证 .....	117
第三节	《高校英语教师语言测评素养量表》的应用 .....	131
<b>第九章</b>	<b>语言测评素养的常态化培养与提升</b> .....	<b>135</b>
第一节	提升语言测评素养的途径与机制 .....	135
第二节	语言测评素养 APP(iLAL)介绍 .....	139
第三节	语言测评素养 APP(iLAL)的应用与展望 .....	146
<b>附录一</b>	<b>教师测评素养问卷节选</b> .....	<b>147</b>
<b>附录二</b>	<b>国际语言测评协会道德准则节选</b> .....	<b>151</b>
<b>参考文献</b>	.....	<b>153</b>
<b>索引</b>	.....	<b>173</b>

# 图目录

图 1-1	语言测评素养的历时发展 .....	001
图 1-2	五元专业发展模型的构成 .....	007
图 1-3	技能+知识+原则模型的构成 .....	009
图 1-4	实践+原则+情境模型的构成 .....	009
图 1-5	语言测评素养掌握程度的高低 .....	010
图 1-6	一线教师所需具备的语言测评素养概貌 .....	012
图 2-1	教师测评素养研究期刊载体分布 .....	018
图 2-2	教师测评素养研究历年发文量和引文量 .....	019
图 2-3	教师测评素养各国发表量分布图 .....	020
图 2-4	教师测评素养研究文献共被引网络聚类知识图谱 .....	021
图 2-5	2005—2018 年间国外教师测评素养研究时间线图 .....	024
图 2-6	外语教师测评素养发展脉络 .....	031
图 3-1	专家视角下语言测评素养构念的修正模型 .....	042
图 3-2	专家视角下英语教师语言测评素养的构成 .....	044
图 5-1	欧洲教育评估标准框架结构 .....	078
图 6-1	英语学科核心素养 .....	086
图 7-1	《中国英语能力等级量表》中写作量表的研制流程 .....	097
图 7-2	《中国英语能力等级量表》的总体结构 .....	108
图 7-3	利益相关方对 CSE 测评素养的掌握程度 .....	112
图 8-1	《高校英语教师语言测评素养量表》开发的流程 .....	118
图 8-2	受试者的群体分布 .....	123
图 8-3	修正后的《高校英语教师语言测评素养量表》结构模型 .....	129
图 8-4	语言测评素养的诊断报告(一) .....	132

图 8-5	语言测评素养的诊断报告(二)	133
图 9-1	iLAL 的扫描安装界面	140
图 9-2	iLAL 的安装流程	141
图 9-3	iLAL 的应用图标	141
图 9-4	iLAL 的主菜单界面	141
图 9-5	《高校英语教师语言测评素养量表》界面	142
图 9-6	根据用户填写数据自动生成的雷达图反馈	143
图 9-7	内嵌于话题栏目的视频资源	144
图 9-8	“测评工具用不用”板块界面	144
图 9-9	“语言测试道德伦理”板块界面	144
图 9-10	讨论区主题列表	145
图 9-11	讨论区主题内容与回复	145

# 表 目 录

表 2-1	教师测评素养各研究主题聚类情况 .....	021
表 2-2	教师测评素养高被引文献信息 .....	023
表 3-1	受访语言测试专家的基本信息 .....	039
表 4-1	语言测试课程用书一览 .....	046
表 4-2	语言测试课程的基本信息 .....	049
表 4-3	语言测试课程的课程目标及内容一览 .....	050
表 4-4	语言测试课程用书与考核方式 .....	053
表 5-1	部分国家和地区的教育或测评标准类文件 .....	061
表 5-2	美国教师专业标准对比 .....	064
表 5-3	《教育与心理测试标准》新旧版对比 .....	067
表 5-4	英国教师专业标准对比 .....	070
表 5-5	澳大利亚《指南》与《标准》的对比 .....	073
表 5-6	新西兰教师教育毕业生标准和“完全注册执业教师” 新标准 .....	076
表 5-7	欧洲标准框架的各维度解析 .....	079
表 5-8	现有语言测评素养量表的有关研究 .....	082
表 6-1	小学阶段英语学科终结性测试构念与题型一览 .....	089
表 7-1	《中国英语能力等级量表》中写作量表参数框架 .....	103
表 8-1	《高校英语教师语言测评素养量表》 .....	119
表 8-2	受试者作答的时长统计 .....	123
表 8-3	受试者作答的信心程度 .....	124
表 8-4	作答数据的 KMO 和 Bartlett 球体测试结果 .....	124
表 8-5	提取因子后的共同性数据(未旋转) .....	124

表 8-6	因子旋转后的结果 .....	126
表 8-7	提取因子的相关矩阵 .....	128
表 8-8	修正模型的拟合度指标 .....	129
表 8-9	修正模型的路径分析结果 .....	130

# 第一章 语言测评素养的基本理论

语言测评素养的研究起源于普通教育学领域,并在语言测试界得到了长足的发展。本章首先从语言测评素养的定义出发,梳理这一概念的本质与内涵,然后逐一列举并解释语言测评素养研究中的四个模型,并对这些模型加以评述。

## 第一节 语言测评素养的定义

“测评素养”一词首次由美国教育评估专家 R. J. 斯蒂金斯(R. J. Stiggins)在 1991 年发表的题为 *Assessment Literacy* 一文中使用,因此,一般认为测评素养主要来源于教育学领域的相关研究。斯蒂金斯(Stiggins, 1995: 240)认为,具备测评素养的人有能力区分高质量和低质量的评估(sound and unsound assessment),并能运用有关知识和理论对学生的学业表现作出合理的推断。最初的测评素养研究涉及教育领域的各方人士,如学校管理者、教师、师范生,更有学者认为准教师必须要“服用测评素养这一剂良方”(a dose of assessment literacy)(Popham, 2006: 86)。

测评素养的定义在之后的研究过程中得到了进一步的发展。比如,帕特诺(Paterno, 2001)将测评素养定义为:对高质量评估基本原理的掌握,包括术语概念、评估方法和技术的开发与应用;对评估质量标准的熟悉程度;对有别于传统测试的替代测试的熟悉程度等。当然,除了通过下定义的方式来对测评素养加以界定外,也可通过列举特征的方式来加以说明。例如,美国博伊西州立大学的学校进步与政策研究中心(Center for School Improvement and Policy Studies)在其官网(<http://csi.boisestate.edu/ah>)中指出具备测评素养的教育者应熟知以下三种高质量的测试和评估行为:①知晓应使用何种评价手段来获

取可靠的学业成就；②通过使用成绩报告、分数、档案袋等方式有效地沟通评价结果；③通过评价行为将学生的潜能最大化，并将学生视为评价过程中不可或缺的主体。

然而，由于研究领域以及不同学科教育的特殊性，语言教育中的测评素养还无法在以上定义中得到充分的阐释。随着语言测试领域研究的不断深入，语言测试领域的学者们（如 Harding & Kremmel, 2016; Inbar-Lourie, 2008; Jeong, 2013; Taylor, 2013）也逐渐意识到测评素养对语言教师的重要意义，并有理由将语言测评素养从普通教育学领域中抽离出来，单独研究语言教育领域中的这一概念。由此，“语言测评素养”（Language Assessment Literacy, LAL）这一概念也应运而生。

语言测评素养的定义在语言测试领域中基本一致。比如，皮尔和哈丁（Pill & Harding, 2013: 382）把语言测评素养定义为“一系列能使个体理解、评估并在一定情况下创造语言测试并分析测试数据的能力集”<sup>①</sup>。值得注意的是，“语言”一词在这一定义中得到了凸显，比前文中斯蒂金斯（Stiggins, 1995）对测评素养的定义更加具体化。

马隆（Malone, 2013: 329）在课堂评估的特定条件下，将测评素养定义为“语言教师对测试理论有关定义的熟知程度，以及将此知识用于一般课堂实践，尤其是用于解决与测评语言相关的问题的能力”<sup>②</sup>。此外，他强调“测评素养也包含掌握试题命制、成绩解释和使用以及试卷评估的能力，同时培养正确理解评估在社会中的作用和功能的能力等”<sup>③</sup>（p. 363）。可见，马隆对测评素养的社会属性（比如，如何在特定的社会背景下组织考试并对成绩进行解释）十分关注。然而，我们也发现，这一定义并未凸显“语言”的特色。换言之，其他领域的教育，如数学、体育也可以运用这一定义。

在众多的定义中，富尔切尔（Fulcher, 2012: 125）对语言测评素养的定义受到较多的肯定和引用，并且随后的研究也有很多是沿着这一定义而展开的（如

---

① 原文为：a repertoire of competencies that enable an individual to understand, evaluate and, in some cases, create language tests and analyze test data.

② 原文为：language instructors' familiarity with testing definitions and the application of this knowledge to classroom practices in general and specifically to issues related to assessing language.

③ 原文为：Nevertheless, acquisition of a range of skills related to test production, test-score interpretation and use, and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within society.

Harding & Kremmel, 2016)。富尔切尔认为,语言测评素养一般包括三个方面,主要涉及“设计、开发、维护或评价大规模标准化考试和课堂测试等所应具备的知识、技能和能力;对考试过程的熟悉程度;了解指导测试实践的原则和概念,如测试伦理、行为准则”。同时他指出,语言测评素养也包括“将有关知识、技能、过程、原则和概念应用到更广泛的历史、社会、政治和哲学框架之中的能力,以便理解评价行为背后的起因,评估测试对社会、机构和个人的作用和影响”<sup>①</sup>。

从以上定义的发展来看,测评素养发展到语言测评素养是一个从普通到具体的演变过程,其中不仅发生了教育领域和场景的具象化,也发生了定义内涵的精细化。因此,国际语言测试学界普遍认为,语言教育者需要具备更多样、更具体的能力和素养来达到语言测评素养的要求。

## 第二节 语言测评素养的发展

第一节内容从定义的角度解释了何为“测评素养”以及何为“语言测评素养”,就本书的研究范畴而言,我们将主要研究方向聚焦后者。本节将在以上有关定义的基础上,以学者们在这一问题上的代表作为主线,厘清语言测评素养研究的发展脉络。

总体而言,现今语言测评素养的发展可通过一个历时途径得以呈现,如图 1-1 所示,首先是以布林德里(Brindley, 2001)发表的作品作为主要标志。他认为,就语言测评素养而言,所有的语言教师应具备以下五个方面的知识:①评价实施的社会环境;②语言能力的定义和描述;③语言测试的命制和评估;④测试在语言教育大纲中的地位;⑤如何将评价付诸实施。

然而,布林德里的分类系统性不强,五个方面的知识内部结构也较为模糊。戴维斯(Davies, 2008)站在更高的角度,对语言测评素养这一概念加以解释。

---

① 原文为: The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals.

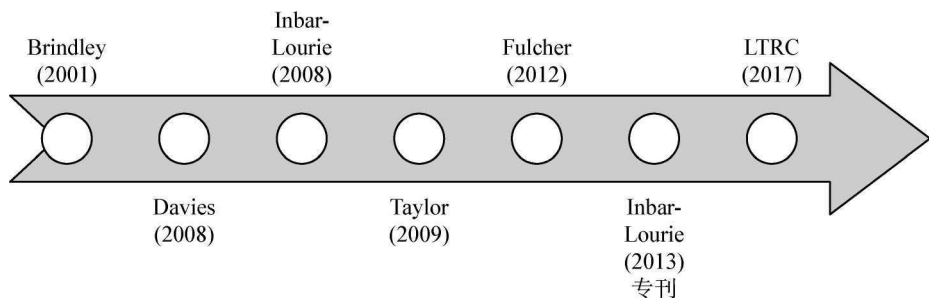


图 1-1 语言测评素养的历时发展

他在论著中对职前和在职教师培训提出三个核心元素,即技能(skills),知识(knowledge)和原则(principles)。其中,技能包括教师对试题命制与编写、统计操作方面的内容;知识包括对语言能力构念以及测量方面的内容;原则主要涉及公平性、考试影响、测试伦理等方面的内容。

因巴尔-劳瑞(Inbar-Lourie, 2008)在戴维斯所提出的三个核心元素基础上进一步发展了语言测评素养所试图解决的问题。她由此提出三个命题,即如何掌握这些技能?需要掌握哪些知识?为何需要遵循这些原则?回答好这三个问题可以有效解决如何培养语言测评素养这一重要命题。此外,2008年对于语言测评素养的发展而言意义非凡,欧洲语言测试者协会(Association of Language Testers in Europe)于2008年年会期间专门研讨了有关语言测评素养研究的话题,引发了语言测试领域新一轮的思考。

之后的发展似乎更倾向于回答“哪些群体需要培养语言测评素养”以及“特定群体需要培养什么样的语言测评素养”这些问题。因此,泰勒(Taylor, 2009: 27)在学者们研究的基础上,进一步提出语言测评素养不仅需要考虑到具备什么技能、知识和原则,还要对“具体教育和社会背景下实施评价的作用和功能有着清晰的认识”。她将语言测评素养落实到不同群体,指出语言测评素养不仅应涉及考试机构和研究人员,也应涉及语言政策的制定者和教育管理者,甚至也包括家长、领导和更为广泛的公众(p. 25)。

富尔切尔(Fulcher, 2012: 125)的研究可谓综合了以往学者们对语言测评素养的论述。他丰富了戴维斯的三大核心元素,进一步将语言测评素养分为实践性知识(practical knowledge)、程序性或理论性知识(procedural/theoretical knowledge)以及社会历史认知(socio-historical understanding)。这三块内容由低到高排列。此外,他还试探性地提出不同群体需要掌握哪些方面的语言测

评素养。

与2008年类似,2013年和2017年也是有关研究成果丰收的两个节点。2013年《语言测试》(*Language Testing*)刊发专刊,由因巴尔-劳瑞作为特约编辑,从语言教师、命题人员、应用语言学研究者、入学管理者、领导决策者五个群体入手,发表了有关语言测评素养的论文。2017年于哥伦比亚首都波哥大举办的国际语言测试大会(Language Testing Research Colloquium, LTRC)则以“跨越利益群体边界的语言测评素养”(Language Assessment Literacy Across Stakeholder Boundaries)为大会主题,集中研讨了语言测试最新的研究成果,将未来语言测试的方向推到了新的高度。学界普遍认为,语言测评素养的发展已从素养的解构转变到了素养的培养对象,从静态素养的概念发展到了动态的变化和全貌性的进步。这些趋势在最新的研究中都得到了充分的体现。

### 第三节 语言测评素养的外延

从语言测评素养的定义及其发展,我们可以管窥到这一概念的内涵与本质,但如何理解语言测评素养的外延?有何规范性的标准文件来规定语言测评素养的有关内容?本节将对此展开讨论。

如前文所述,语言测评素养是研究者们(Brindley, 2001; Brown & Bailey, 2008; Coombe, Troudi, & Al-Hamly, 2012; Fulcher, 2012; Harding & Kremmel, 2016; Inbar-Lourie, 2013等)在教育领域逐渐发展起来的研究。因此,语言测评素养的本质应扎根于教育学,并对教育学中的有关理论和标准具有反向的指导意义。这种指导的直接产物之一就是美国教师联合会(American Federation of Teachers, AFT)、全美教育测量委员会(National Council on Measurement in Education, NCME)和全美教育协会(National Education Association, NEA)于1990年所颁布的《学生教育评价中的教师能力标准》(*The Standards for Teacher Competence in Educational Assessment of Students*)将测评素养纳入其中,并提出七条标准,具体如下:

标准1——教师应熟练选择适合教学决策的评估方法。

标准2——教师应擅长开发适合教学决策的评估方法。

标准 3——教师应熟练管理、评分和解释外部产生的和教师产生的评估方法的结果。

标准 4——教师在对个别学生做出决策、规划教学、开发课程和改进学校管理时,应熟练使用评估结果。

标准 5——教师应熟练使用学生评估制定有效的学生评分程序。

标准 6——教师应熟练地将评估结果传达给学生、家长、其他非专业受众和教育工作者。

标准 7——教师应熟练识别不道德、非法和其他不当的评估方法和评估信息。

从标准的内容中可以发现,标准 1 和标准 2 主要聚焦教师应熟知的评价方法和手段;标准 3 规定了教师对施考、成绩解释的有关要求;标准 4 强调了考试和成绩的使用;标准 5 说明了评价过程中学生参与的重要性;标准 6 则进一步强调了考试结果的沟通和透明度;标准 7 落实到了评价的伦理道德问题上。这些标准成为今后测评素养研究的主要标尺(Brookhart, 2001),也是衡量语言测评素养高低工具(量表)开发的重要依据(详见本书第八章)。

#### 第四节 语言测评素养的模型

本章第二节呈现了语言测评素养研究的基本发展脉络,这一过程不仅是一条时间轴的顺序,也是语言测评素养模型的发展顺序。本节将依次具体介绍布林德里(Brindley, 2001)的五元专业发展模型(five-component professional development program model)、戴维斯(Davies, 2008)和因巴尔-劳瑞(Inbar-Lourie, 2008)的技能+知识+原则模型(skills, knowledge and principles model)、富尔切尔(Fulcher, 2012)的实践+原则+情景模型(practices, principles and contexts model)以及皮尔与哈丁(Pill & Harding, 2013)和泰勒(Taylor, 2013)的语言测评素养利益相关方的概貌模型(LAL stakeholder profile model)。本节的最后将对这些模型加以评述。

##### 模型一：五元专业发展模型

这一模型的源泉仍是第三节介绍的语言测评素养的外延,即美国所颁布的

《学生教育评价中的教师能力标准》。布林德里(Brindley, 2001)在充分评价《学生教育评价中的教师能力标准》的基础上认为,除了掌握的核心素养外,不同的个体对素养的掌握程度以及他们对考试或评价的参与度不尽相同。由此,布林德里(Brindley, 2001)所提出的五元专业发展模型顾名思义由五块内容组成,其中包含两个核心单元和三个选修单元,如图 1-2 所示:



图 1-2 五元专业发展模型的构成

在布林德里的五元专业发展模型中,核心单元之一是知晓本地考试与评估在社会、教育、政治以及伦理方面的特征。这一单元的意义在于在不同的社会文化背景下,人们对考试与评估的认知视角或价值观有所不同。比如,在有些社会文化背景下,相较于终结性评价而言,语言教育者可能对形成性评价更为重视。此外,由于社会文化等因素影响着人们对考试效度的看法,因此教师应该具备一定的能力,如在设计考试、报道分数时有意识地兼顾到各项社会因素。同样,政治和伦理方面的素养也同等重要。比如,通过一些手段来区分考生,或是突出某些社会特权阶层都将给教育考试带来极大的弊端(Shohamy, 2001)。又如,如果教师不懂得如何在考试过程中让弱势群体(如残疾人士)避免受到歧视,那么此项教育考试的伦理程度也随之降低。另一核心单元则涉及对语言能力水平的界定和描述。布林德里(Brindley, 2001)认为,掌握有关语言能力模型(如 Bachman, 1990; Bachman & Palmer, 1996; Canale & Swain, 1980)并指出批判性的评价是语言测评素养的主要内容。

就选修单元而言,布林德里(Brindley, 2001)认为,语言教师首先应具备开发和评估语言考试的能力,并初步掌握有关统计分析的内容。另一选修单元涉及标准参照考试的有关内容、替代性评价(如档案袋、学习日志、学生自评)的有

关内容以及如何语言教学大纲中融入这些评价。最后一个选修单元提供了应用策略上的路线图,对以上所有单元在实践过程中可能发生的困难给予帮助。

这一语言测评素养模型为今后的研究奠定了一定的基础,特别是核心单元强调语言教育者需要掌握本地语言测评的背景和现状,这与语言测试的社会化属性息息相关(McNamara & Roever, 2006; Shohamy, 2001)。因巴尔-劳瑞(Inbar-Lourie, 2008)在推出其新模型(即技能+知识+原则模型)时,对五元专业发展模型做出如下评述:五元专业发展模型中的核心单元一对应新模型中的 why,即为何要实施评价;核心单元二则对应其新模型中的 what,即语言能力评价的本质和特征;其余的选修单元则对应 how,即评价的手段与方法。

当然,也有学者对这一模型持不同的意见。比如,因巴尔-劳瑞(Inbar-Lourie, 2008)发现这一模型虽然强调了语言测试的社会化属性,但是对语言测试如何在教育决策中起到积极作用似乎并未提及。此外,不少学者认为这一模型中的核心和选修单元从分类角度而言本身也存在瑕疵。哈丁和克雷姆(Harding & Kremmel, 2016: 417)认为:这一模型中的五块内容都是核心,对语言教师的测评素养而言均为“必修课程”。

#### 模型二:技能+知识+原则模型

这一模型是集合了因巴尔-劳瑞(Inbar-Lourie, 2008)和戴维斯(Davies, 2008)的有关论述。后者从历时的角度对2007年之前出版的有关语言测试教科书进行了综述,其中最早的教科书可追溯到拉多(Lado, 1961)编写的《语言测试》(*Language Testing*),最新的为富尔切尔和戴维森(Fulcher & Davidson, 2017)出版的《语言测试与评估:高级资源手册》(*Language Testing and Assessment: An Advanced Resource Book*)。在综述中,戴维斯(Davies, 2008)提炼出语言测试教科书的一般脉络,即涉及“技能”“知识”和“原则”三个方面的内容。戴维斯(Davies, 2008)认为,技能是指恰当的评价方法和工具等,如试题命制、统计分析、成绩报告;知识是指教育测量和语言能力水平方面的基础以及本地化的现实因素等;原则则是指评价的使用、影响、公平性和伦理方面的内容等。这一分类与因巴尔-劳瑞同年发表的论文中对语言测评素养的三大问题,即 how, what 和 why 可谓不谋而合,如图 1-3 所示:

#### 模型三:实践+原则+情境模型

富尔切尔(Fulcher, 2012: 125)在以上技能+知识+原则模型的基础上,通