



国际中文教师语言测评素养： 基于语言测评知识、实践和反思的调查

张艳莉 王肖竹

摘要:语言测评素养是语言教师合理认识、设计测评活动和运用测评结果所必需的专业素养。基于语言测评知识、实践和反思的角度调查国际中文教师语言测评素养,发现:国际中文教师的语言测评知识处于中等水平,教师对语言测评相关的概念和原则有一定了解,但存在理解不充分以及误解的情况,参加测评培训对语言测评知识水平有提升作用;教师在教学实践中注重形成性评价,也存在一定的高风险测试导向;教学经验和测评培训因素均对教师语言测评素养自评和提升需求产生积极影响,教师对语言测评素养的提升持积极态度,需要语言测评相关培训提供具体实践性的指导。

关键词:语言测评素养;国际中文教师;语言测评知识;教师专业发展

DOI: 10.13734/j.cnki.1000-5315.2023.02.017

收稿日期: 2022-10-09

基金项目: 本文系汉考国际科研基金重点项目“汉语水平考试认知效度研究”(CTI2021A01)的阶段性研究成果。

作者简介: 张艳莉,女,山东曲阜人,上海外国语大学国际文化交流学院教授,E-mail: zhangyanli@shisu.edu.cn;

王肖竹,女,山东威海人,北京语言大学国际学生教育政策与评价研究院博士研究生。

一 引言

测试与评价是语言学习中不可缺少的一部分,要判断测评是否有效地反映学生的语言水平和学习情况,解读测评信息以指导教学或决策,需要与测评相关的各方(学生、教师、学校管理者等)具有一定的语言测评素养。在语言测试领域,对测试相关方语言测评素养的关注度持续提升^①。其中,语言教师的测评素养最受关注,因其承担着课堂测评活动设计和实施者的角色,语言测评素养对教师测评活动的开展及教、学、评一体化发展有重要作用。因此,语言测评素养也是教师专业发展的一环,缺乏测评素养意味着专业性的丧失^②。

在国内,国际中文教师主要对来华留学生进行汉语作为第二语言的教学,其语言测评素养较少受到关注。一般而言,教师的测评理念和方式与其所接受的教育十分相关^③,这种“沿袭”的做法不利于测评的创新和发展。国际中文教师面对的学生来自不同的教育文化环境,其语言测评素养需要兼顾高风险考试和课堂测评,并用学生可以接受的方式设计和实施测评活动。目前,对国际中文教师能力的认定标准均涉及测评素

^①Geoff Brindley, "Language Assessment and Professional Development," in *Experimenting with Uncertainty: Essays in Honour of Alan Davies*, eds. C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, K. O'Loughlin (Cambridge: Cambridge University Press, 2001), 126-136; Lynda Taylor, "Communicating the Theory, Practice and Principles of Language Testing to Test Stakeholders: Some Reflections," *Language Testing* 30, no. 3 (July 2013): 403-412; Christine Coombe, Hossein Vafadar, Hassan Mohebbi, "Language Assessment Literacy: What Do We Need to Learn, Unlearn, and Relearn?," *Language Testing in Asia* 10, no. 3 (May 2020).

^②W. James Popham, "Why Assessment Illiteracy is Professional Suicide," *Educational Leadership* 62, no. 1 (September 2004): 82-83.

^③Karin Vogt, Dina Tsagari, "Assessment Literacy of Foreign Language Teachers: Findings of a European Study," *Language Assessment Quarterly* 11, no. 4 (November 2014): 374-402; Ricky Lam, "Teacher Assessment Literacy: Surveying Knowledge, Conceptions and Practices of Classroom-based Writing Assessment in Hong Kong," *System* 81, (April 2019): 78-89.

养,如从“测试与评估”^①、“学习评估和反馈”^②方面对教师所具备的测评能力提出要求,相关的教师培训项目应当包含语言测评课程。现有的针对国际中文教师的语言测评课程效果如何?教师对学生的测评是怎样进行的?目前尚缺乏对这些问题的探讨。

作为教学与测试的语言能力水平依据,《国际中文教育中文水平等级标准》^③(以下简称《等级标准》)的发布和使用要求教师据此评估和追踪学生语言发展。此外,自受新冠疫情影响以来,在线教学模式得到了全面深入发展^④,语言测评的开展形式也发生了相应变化。无论是指导标准的规范化还是语言测评形式的转变,都对国际中文教师的语言测评素养提出了更为具体的要求,教师需要对语言测评的原则和方法有系统的知识基础,并在此指导下选择、改编或设计适合课程教学目标的测评活动和考试来促进教学和学习。在此背景下,本研究尝试从知识、实践和反思的角度对国际中文教师的语言测评素养进行调查,并探索提升的方向。

二 文献回顾

(一)语言测评素养内涵与模型

测评素养是一个多层面的概念,最初由教育学界提出,由于教育问责制(Accountability)以及促学评价理念(Assessment for learning)的发展,要求教师及其他测评相关方具有正确使用测评结果的能力^⑤。在国际上,语言测试作为测量语言能力水平的工具,与移民和留学等政策息息相关,其应用得到扩展,影响力也在增大^⑥。随着语言测试专业化发展和学科知识体系的形成,语言测评素养作为测评素养的子集被提出^⑦,其构念(Construct)和内容融合了测评素养的基本内涵和语言测试学科的特点。为了探寻语言测评素养所涉及的内容,学者对语言测试指导性书籍进行了分析^⑧,对以提升语言测评素养为目的的培训课程进行了调研^⑨,并了解了教师学习语言测评知识的需求^⑩。在具体内容的基础上,形成了“技能+知识+原则”^⑪、“目的+内容+方式”^⑫、“环境+原则+实践”^⑬等语言测评素养的结构模型。在已有研究基础上,金艳根据我国外语教育领域的主要测评活动分类(大规模考试、学业测试和课堂评价),提出了外语教师测评素养的模型,包含环境、原则、实践三个层面,以及外语教师在不同类型测评活动中语言测评素养的要求^⑭。综合已有的定义,本研究将教师语言测评素养的概念界定为:语言教师对于大规模标准化考试和课堂测评的基本概念和指导原则的了解以及设计、实施和利用测评活动的知识、技能和能力,并能从社会文化和政策环境的角度理解测试的性质及其影响。

近年来的研究多从社会文化角度对语言测评素养进行解读。如 Xu & Brown 在回顾百篇测评素养研究文献的基础上,提出了实践中的教师测评素养模型(Teacher Assessment Literacy in Practice, TALiP),包含“知识基础”、“教师测评观念”、“体制和社会文化环境”、“实践中的测评素养”、“教师学习”、“教师测评者身

①国家汉语国际推广领导小组办公室《国际汉语教师标准》,外语教学与研究出版社2007年版;孔子学院总部、国家汉办《国际汉语教师标准》,外语教学与研究出版社2015年版。

②世界汉语教学学会《国际中文教师专业能力标准》,2022年。

③教育部、国家语言文字工作委员会《国际中文教育中文水平等级标准》,2021年。

④亓海峰、丁安琪《海外汉语教师在线教学现状调查分析》,《天津师范大学学报(社会科学版)》2021年第5期,第42-47页。

⑤Richard J. Stiggins, "Assessment Literacy," *The Phi Delta Kappan* 72, no. 7 (March 1991): 534-539; Christopher DeLuca, "Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education," *Action in Teacher Education* 34, no. 5-6 (December 2012): 576-591.

⑥Glenn Fulcher, "Assessment Literacy for the Language Classroom," *Language Assessment Quarterly* 9, no. 2 (May 2012): 113-132.

⑦林敦来、高森《教师评估素养:理论与实践》,《外语教学理论与实践》2011年第4期,第29-37页。

⑧Alan Davies, "Textbook Trends in Teaching Language Testing," *Language Testing* 25, no. 3 (July 2008): 327-347.

⑨James D. Brown, Kathleen M. Bailey, "Language Testing Courses: What are They in 2007?," *Language Testing* 25, no. 3 (July 2008): 349-383.

⑩Glenn Fulcher, "Assessment Literacy for the Language Classroom," 113-132.

⑪Alan Davies, "Textbook Trends in Teaching Language Testing," 327-347.

⑫Ofra Inbar-Lourie, "Constructing a Language Assessment Knowledge Base: A Focus on Language Assessment Courses," *Language Testing* 25, no. 3 (July 2008): 385-402.

⑬Glenn Fulcher, "Assessment Literacy for the Language Classroom," 113-132.

⑭金艳《外语教师评价素养发展:理论框架和路径探索》,《外语教育研究前沿》2018年第2期,第65-72、93页。

份”六个部分^①。该模型反映了测评素养积累、实践、反思的动态过程,并强调测评知识基础并不能直接成为测评素养,需要通过教师理解、实践、进一步学习和反思才能完成知识到素养的转化,且教师的测评实践通常是在多种因素的影响和权衡下实现的。Pastore & Andrade 通过专家咨询法归纳了包含“概念知识”、“行动”和“社会情感”三个维度的测评素养模型^②。潘鸣威通过专家访谈更新的语言测评素养模型也将教师情感和态度作为一个外延^③。综合以上框架内容可以看出,教师需要一定的语言测评知识基础,其内容来源于语言测试学科范围,并与社会文化、教育政策与教学环境相协调,从而在实践中具备相应的技能和素质^④。

(二)教师语言测评素养研究现状

已有的教师语言测评素养研究内容和方法主要集中于以下五个方面:(1)调查工具的开发和验证,形式有基于内容的选择题项^⑤、自评量表^⑥、判断题项^⑦、选择题+自评量表^⑧等;(2)量化评估教师测评素养的不足和培训需求^⑨,涉及群体包括国内中小学教师、欧洲外语教师、高校英语教师等,并探讨了影响测评素养水平及发展的关键因素,如测评培训、教学经验、环境因素等^⑩;(3)教师与专家对语言测评素养理解的差异,认为语言测评素养培训中应考虑到教师的理解特点和实践需求^⑪;(4)教师的语言测评实践,通过分析测评活动设计^⑫、命题质量^⑬、课堂观察^⑭、访谈^⑮等内容总结出教师在实践中体现出的语言测评素养;(5)教师观

① Yueting Xu, Gavin T. L. Brown, “Teacher Assessment Literacy in Practice: A Reconceptualization,” *Teaching and Teacher Education* 58, (August 2016): 149-162.

② Serafina Pastore, Heidi L. Andrade, “Teacher Assessment Literacy: A Three-dimensional Model,” *Teaching and Teacher Education* 84, (August 2019): 128-138.

③ 潘鸣威《外语教师语言测评素养再探——基于对语言测试专家的访谈》,《中国考试》2020年第7期,第34—41页。

④ Ofra Inbar-Lourie, “Constructing a Language Assessment Knowledge Base: A Focus on Language Assessment Courses,” 385-402.

⑤ Barbara S. Plake, James C. Impara, Jennifer J. Fager, “Assessment Competencies of Teachers: A National Survey,” *Educational Measurement: Issues and Practice* 12, no. 4 (December 1993): 10-12.

⑥ 肖阳田《中学英语教师测评素养自评量表开发与验证》,上海外国语大学2020年博士学位论文; Fateme Nikmard, Zohre Mohamadi Zenouzagh, “Designing and Validating a Potential Assessment Inventory for Assessing ELTs’ Assessment Literacy,” *Language Testing in Asia* 10, no. 8 (July 2020); Benjamin Kremmel, Luke Harding, “Towards a Comprehensive, Empirical Model of Language Assessment Literacy Across Stakeholder Groups: Developing the Language Assessment Literacy Survey,” *Language Assessment Quarterly* 17, no. 1 (2020): 100-120; Zia Tajeddin, Mohammad Khatib, Mohsen Mahdavi, “Critical Language Assessment Literacy of EFL Teachers: Scale Construction and Validation,” *Language Testing* 39, no. 4 (October 2022): 649-678.

⑦ Elçin Ölmezler-Öztürk, Belgin Aydin, “Toward Measuring Language Teachers’ Assessment Knowledge: Development and Validation of Language Assessment Knowledge Scale (LAKS),” *Language Testing in Asia* 8, no. 20 (2018).

⑧ Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe-McEwan, Ulemu Luhanga, “Teachers’ Approaches to Classroom Assessment: A Large-scale Survey,” *Assessment in Education: Principles, Policy & Practice* 25, no. 4 (2018): 355-375.

⑨ 郑东辉《中小学教师评价素养状况:来自Z省的报告》,《全球教育展望》2010年第2期,第31-36、42页; Karin Vogt, Dina Tsagari, “Assessment Literacy of Foreign Language Teachers: Findings of a European Study,” 374-402; 江进林《高校英语教师测评素养现状及其影响因素研究》,《外语界》2019年第6期,第18—26页。

⑩ Craig A. Mertler, “Secondary Teachers’ Assessment Literacy: Does Classroom Experience Make a Difference?,” *American Secondary Education* 33, no. 2 (Spring 2005): 76-92; Ling Gan, Ricky Lam, “Understanding University English Instructors’ Assessment Training Needs in the Chinese Context,” *Language Testing in Asia* 10, no. 11 (September 2020); Xun Yan, Cong Zhang, Jason Jinsong Fan, “‘Assessment Knowledge is Important, but ...’: How Contextual and Experiential Factors Mediate Assessment Practice and Training Needs of Language Teachers,” *System* 74, (June 2018): 158-168.

⑪ Heejeong Jeong, “Defining Assessment Literacy: Is it Different for Language Testers and Non-language Testers?,” *Language Testing* 30, no. 3 (July 2013): 345-362; Margaret E. Malone, “The Essentials of Assessment Literacy: Contrasts between Testers and Users,” *Language Testing* 30, no. 3 (July 2013): 329-344.

⑫ Kim Koh, Lydia E. Carol-Ann Burke, Alan Luke, Wengao Gong, Charlene Tan, “Developing the Assessment Literacy of Teachers in Chinese Language Classrooms: A Focus on Assessment Task Design,” *Language Teaching Research* 22, no. 3 (May 2018): 264-288.

⑬ Cong Zhang, Xun Yan, “Assessment Literacy of Secondary EFL Teachers: Evidence from a Regional EFL Test,” *Chinese Journal of Applied Linguistics* 41, no. 1 (2018): 25-46, 3.

⑭ Ricky Lam, “Teacher Assessment Literacy: Surveying Knowledge, Conceptions and Practices of Classroom-based Writing Assessment in Hong Kong,” 78-89.

⑮ Xun Yan, Cong Zhang, Jason Jinsong Fan, “‘Assessment Knowledge is Important, but ...’: How Contextual and Experiential Factors Mediate Assessment Practice and Training Needs of Language Teachers,” 158-168.

念、反思等社会文化要素,通过混合研究方法反映教师对语言测评活动的认识与态度^①,理解教师观念对于相关政策和培训课程的实施有重要意义^②,教师反思则是发挥教师能动性、在现有实践基础上寻求提升的有效方式^③。

目前以国际中文教师为对象的语言测评素养研究较少,鉴于研究环境的差异性及其研究内容对国际中文教师发展的意义,有必要探讨高校国际中文教师语言测评素养的现状,以更好地在教师职前培养阶段和专业发展上提供帮助。基于以上语言测评素养的理论模型和实证研究,本研究希望回答以下问题:(1)国际中文教师语言测评知识水平如何?(2)国际中文教师的语言测评实践有何特点?(3)从自评和提升需求的角度看,国际中文教师对其语言测评素养的反思情况如何?

三 研究设计

(一)研究对象

本研究的对象为在高校从事国际中文教学工作的教师,通过便利抽样发放问卷,共得到有效答卷 119 份。高校所在地分布于上海、山东、辽宁、甘肃、吉林、北京、浙江、广东等 16 个省市,教师的教龄分布为 0—2 年(32.77%),3—10 年(32.77%),10 年以上(34.45%)。本研究将教龄为 0—2 年界定为新手教师(39 人)^④,大多为正在接受或刚刚完成汉语国际教育硕士项目的学习;教龄为 3—10 年的教师与 10 年以上的教师分别为熟练型(39 人)、专家型教师(41 人),一般为高校在职教师,教学经验比较丰富。74 名(62.18%)教师参加过测评培训,其中近三年参加过的有 52 人(46.22%),另外 45 人(37.82%)未参加过测评培训。共有 8 名教师参加了后续访谈,其中新手教师 4 人,熟练型教师 1 人,专家型教师 3 人。

表 1 调查对象的教学经验和培训经历分布

参加培训情况/教学经验	新手教师	熟练型教师	专家型教师	总计(参加测评培训情况)
未参加过测评培训	16	13	16	45
参加过测评培训且在三年前	3	7	12	22
参加过测评培训且在三年内	20	19	13	52
总计(教学经验分类)	39	39	41	119

(二)研究框架及工具

为了从多元的角度反映教师语言测评素养,本研究采用 Xu & Brown 的 TALiP 模型作为融合理论知识与教师实践的宏观理论框架^⑤。研究主要围绕国际中文教师的“语言测评知识”、“语言测评实践”、“语言测评反思”三方面进行探讨。

本研究的分析和解释框架如图 1,通过问卷调查和访谈的方法收集数据,共使用“语言测评知识问卷”和“语言测评素养评价量表”两份问卷。在制定“语言测评知识问卷”时,首先通过筛选和整合已有文献中语言

^①Angela Scarino, “Language Assessment Literacy as Self-awareness: Understanding the Role of Interpretation in Assessment and in Teacher Learning,” *Language Testing* 30, no. 3 (July 2013): 309-327; Deborah Crusan, Lia Plakans, Atta Gebriel, “Writing Assessment Literacy: Surveying Second Language Teachers’ Knowledge, Beliefs, and Practices,” *Assessing Writing* 28, (April 2016): 43-56; Ricky Lam, “Teacher Assessment Literacy: Surveying Knowledge, Conceptions and Practices of Classroom-based Writing Assessment in Hong Kong,” 78-89; Tziona Levi, Ofra Inbar-Lourie, “Assessment Literacy or Language Assessment Literacy: Learning from the Teachers,” *Language Assessment Quarterly* 17, no. 2 (2020): 168-182.

^②Gavin T. L. Brown, “Teachers’ Conceptions of Assessment: Implications for Policy and Professional Development,” *Assessment in Education: Principles, Policy & Practice* 11, no. 3 (2004): 301-318.

^③Esmat Babaii, Fatemeh Asadnia, “A Long Walk to Language Assessment Literacy: EFL Teachers’ Reflection on Language Assessment Research and Practice,” *Reflective Practice* 20, no. 6 (2019): 745-760.

^④研究对象的划分主要根据每个分组中国际中文教师经历的同质性以及每组人数的平衡性。

^⑤Yueting Xu, Gavin T. L. Brown, “Teacher Assessment Literacy in Practice: A Reconceptualization,” 149-162.

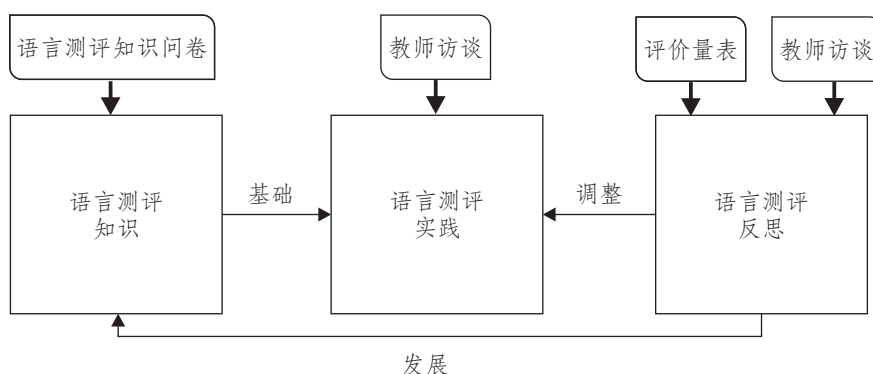


图1 改编与整合的研究设计框架

测评素养包含的内容^①、教师课堂测评标准^②,并考虑本研究环境,确定问卷的内容结构,然后参考已有的测量工具的问题设置^③进行改编,主要从情境和语言方面对问题进行修改,使问题情境符合汉语作为第二语言教学的实际情况,考虑文化差异因素,并结合课程评价体系、线上测试的实施情况进行情境设定;在测评的语言对象上,考虑汉语的语言标准、语言要素、语言技能和语言使用等对内容进行调整,例如在环境层面,主要考查教师对汉语教学大纲在教学中的指导作用及对《等级标准》发布背景和意义的了解,但对语言测评基本概念和原则相关的题目基本不作改变。制定好的问卷题目及答案经过3名具有语言测试专业知识的专家审核和修改,此后经过小范围试测确认问卷达到满意的信度,并根据试测结果修订了问卷。最终问卷涵盖环境、原则、实践3个层面,包含19个主题的内容(见表2),由26个选择题构成,以0—1计分(对多个答案的题目每个选项单独计分),共得到77个计分项用于描述性统计和方差分析。“语言测评素养评价量表”包含自评量表和提升需求量表,根据DeLuca等课堂测评方法调查问卷(ACAI)^④进行改编,为Likert五度量表,共24个题项(见表3、4)。在问卷发放之后,对部分教师进行了半结构化访谈,访谈内容主要包括教师对语言测评活动的认识和态度等,教师所接触和实施的语言测评活动(课程评估体系、具体测评方式和实施情况等),以及对语言测评进行的反思,包括语言测评素养的发展与提升需求等。

(三)数据收集和分析

本研究的数据包括定量和定性两部分。定量数据通过问卷网发放,使用SPSS 23.0对结果进行统计分析,两份问卷整卷Cronbach's α 系数为0.893。其中,“语言测评知识问卷”总体Cronbach's α 系数为0.86,其中环境部分为0.65,原则部分0.70,实践部分0.80,KMO取值为0.838,Bartlett's球形检验显著($p < 0.000$),说明问卷有较好的信度和结构效度;量表部分整体Cronbach's α 系数为0.926,“语言测评素养自评量表”Cronbach's α 系数为0.922,KMO取值为0.909,Bartlett's球形检验显著($p < 0.000$),通过探索性因子分析,使用主成分分析法,根据最大方差法旋转后得到的两个因子,分别被命名为“测评设计及质量”、“测评准备及反馈”,因子载荷分别为33.217%,30.050%,内部一致性信度分别为0.89,0.85;“语言测评素养提升需求量表”Cronbach's α 系数为0.904,KMO取值为0.869,Bartlett's球形检验显著($p < 0.001$),用同样的方法进行探索性因子分析,得到“测评设计与实施”、“语言测评与反馈”、“测评结果利用”三个因子,因子载荷分别为26.239%,22.993%,20.235%,内部一致性信度分别为0.86,0.83,0.80。为了探究教学经验(新手教师、熟练型教师、专家型教师)与测评培训因素(参加过测评培训、未参加过测评培训)对两份问卷得分产生的影响,进

①Glenn Fulcher, "Assessment Literacy for the Language Classroom," 113-132;金艳《外语教师评价素养发展:理论框架和路径探索》,《外语教育研究前沿》2018年第2期,第65-72、93页。

②Joint Committee for Standards on Educational Evaluation, "Classroom Assessment Standards for PK-12 Teachers," 2015.

③Barbara S. Plake, James C. Impara, Jennifer J. Fager, "Assessment Competencies of Teachers: A National Survey," 10-12; Christopher DeLuca, Danielle LaPointe-McEwan, Ulemu Luhanga, "Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy," *Educational Assessment* 21, no. 4 (September 2016): 248-266.

④Christopher DeLuca, Danielle LaPointe-McEwan, Ulemu Luhanga, "Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy," 248-266.

行 3(3 组)×2(2 组)的混合设计的方差分析。

表 2 语言测评知识问卷内容及结果

层面	内容			平均分	标准差
1.环境	1.1	语言能力标准、大纲和测评目标	Q1, Q5	0.62	0.18
	1.2	《等级标准》研制目的和意义	Q2	0.70	0.31
	1.3	《等级标准》研发理念	Q3	0.54	0.30
	1.4	《等级标准》创新概念	Q4	0.54	0.26
2.原则	2.1	三种测评理念的特点	Q6	0.50	0.27
	2.2	信度和效度的概念	Q7	0.76	0.23
	2.3	汉语测评的效度	Q8, Q9, Q12	0.55	0.30
	2.4	表现型评价的原则	Q10	0.67	0.26
	2.5	终结性评价的特点	Q11	0.71	0.27
3.实践	3.1	线上测试设计和特点	Q14	0.58	0.22
	3.2	课堂评价的目的和做法	Q13, Q15	0.75	0.22
	3.3	高风险考试备考的做法	Q16	0.77	0.22
	3.4	写作命题的做法	Q17	0.62	0.22
	3.5	四类题型的特点	Q19	0.65	0.30
	3.6	口语和写作评分	Q18, Q20	0.57	0.23
	3.7	测评结果分析和利用	Q21, Q22, Q23	0.54	0.32
	3.8	提升效度的做法	Q24	0.51	0.27
	3.9	提升信度的做法	Q25	0.63	0.24
	3.10	提升公平性的做法	Q26	0.64	0.25

定性数据通过半结构化访谈进行收集,其中 2 名教师的访谈在线下当面进行,另外 6 名教师通过电话访谈,访谈时间设定为 20—30 分钟。访谈内容经过文字转写后使用 Nvivo 12.0 进行编码与分析,内容主题包括教师语言测评观念、语言测评实践和语言测评素养的发展三个方面。

表 3 语言测评素养自评量表内容及结果

序号	自评量表题项	平均分	标准差
1	我的测评活动与教学大纲和课程要求契合。	4.07	0.79
2	我能运用已有的语言能力标准,对学生的语言能力做出合理的评价。	3.96	0.87
3	我能根据汉语作为第二语言学习的特点,选择和设计合适的测评方式、方法。	3.94	0.85
4	我能通过测评活动中的信息和经验提高教学能力。	4.22	0.77
5	我的成绩报告是基于充分考察证明的结果,且能反映学生对于课程期望目标的学习成果。	3.97	0.72
6	我让学生监控自己的学习,并运用测评结果制定个性化学习计划和学习方法。	3.53	0.95
7	我能够清晰表达自己的测评理念,认识到其与教学大纲的契合之处与不匹配处。	3.80	0.88

8	我的测评有明确的目的(诊断性、形成性、终结性等),能够支持教学活动。	3.97	0.80
9	在测评之前我为学生提供了足够的资源、时间和学习机会去准备。	4.08	0.81
10	我能够在合适的时机与学生(或其他相关方)等交流测评目的和结果。	3.97	0.88
11	我能运用各种方法来分析测评结果,包括单个学生和班级层面的信息。	3.82	0.86
12	我能提供及时、有效的反馈促进学生学习。	4.10	0.85
注:1—7属于因子一(测评设计及质量),8—12属于因子二(测评准备及反馈)			

表4 语言测评素养提升需求量表内容及结果

序号	提升需求量表题项	平均分	标准差
1	根据教学目标,制定合理的测评目标,并选择合适功能的测评(如诊断性测试、形成性测试、终结性测试等)。	4.29	0.75
2	设计测评活动(依据现有的理论、原则和实践准则)。	4.25	0.83
3	实施和管理测评活动(依据现有的理论、原则和实践准则)。	4.08	0.80
4	为测试评分(依据现有的理论、原则和实践准则)。	4.07	0.84
5	运用已有的语言能力标准,对学生的语言能力做出合理的评价。	4.05	0.94
6	了解汉语教学与测试的特点。	4.00	0.97
7	在测评理论指导下解读和使用测评结果信息。	4.00	1.11
8	与学生、家长、学校等利益相关方交流测评目的、过程和结果。	4.00	0.88
9	为学生营造平等的测评环境,关注到文化等方面差异对公平性的影响。	3.95	0.87
10	保密成绩、保护学生的隐私。	3.88	0.87
11	将形成性评价融入教学指导中。	3.87	0.92
12	分析测评信息并用来指导下一步教学、促进学生学习。	3.61	1.01
注:1—4属于因子一(测评设计与实施),5—9属于因子二(语言测评与反馈),10—12属于因子三(测评结果利用)			

四 结果与讨论

(一)研究结果

1.语言测评知识

语言测评知识问卷的结果显示国际中文教师平均正确作答62.63%的选项($M=48.23$, $SD=10.30$, $SE=0.94$, $Min=20$, $Max=74$),分数分布见图2。其中环境层面59.91%,原则层面64.48%,实践层面62.90%。多项选择题和单项选择题的作答情况分别如表5、6,其中单项选择题正确比例平均为53.32%,说明对教师来说题目难度较大,多项选择题全部选对的比例平均为20.76%,部分选对的比例为48.34%,多数教师仅部分掌握题目内容(少选)。综合以上结果可以看出,国际中文教师具有中等水平的语言测评知识,即对于问卷中涉及的语言测评概念和原则,教师有一定的了解,但了解不充分且存在误解的情况。

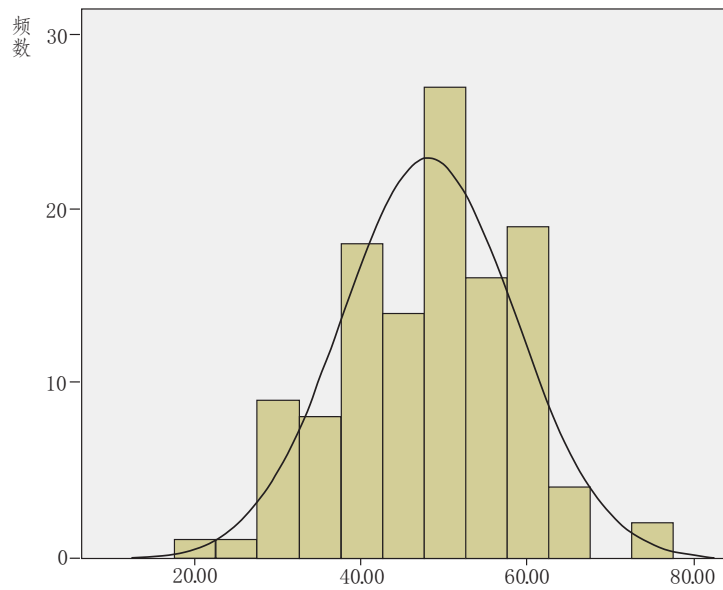


图 2 语言测评知识问卷分数分布直方图

表 5 语言测评知识问卷单项选择题作答情况

单项选择题号	Q5	Q8	Q9	Q12	Q13	Q18	Q21	Q22	Q23
正确率	0.28	0.49	0.59	0.59	0.64	0.61	0.47	0.56	0.58
错选率	0.72	0.51	0.41	0.41	0.36	0.39	0.53	0.44	0.42

表 6 语言测评知识问卷多项选择题作答情况

多项选择题号	正确率	少选率	错选率
Q1	0.15	0.50	0.34
Q2	0.38	0.38	0.24
Q3	0.12	0.25	0.63
Q4	0.04	0.13	0.82
Q6	0.09	0.47	0.44
Q7	0.34	0.48	0.18
Q10	0.23	0.36	0.41
Q11	0.36	0.50	0.14
Q14	0.08	0.76	0.08
Q15	0.40	0.54	0.06
Q16	0.34	0.35	0.30
Q17	0.13	0.66	0.21
Q19	0.30	0.70	0.00
Q20	0.11	0.19	0.70

Q24	0.07	0.56	0.37
Q25	0.17	0.71	0.12
Q26	0.22	0.66	0.12

教学经验对语言测评知识问卷总分和各层面上没有产生显著影响, $F(2, 116) = 0.406, p = 0.667, \eta p^2 = 0.007$; 测评培训因素对语言测评知识问卷的得分有显著影响, $F(1, 117) = 6.344, p = 0.013, \eta p^2 = 0.053$ 。主要体现在环境层面 $F(1, 117) = 6.640, p = 0.011, \eta p^2 = 0.055$, 和实践层面 $F(1, 117) = 5.049, p = 0.027, \eta p^2 = 0.043$ 。关于教学经验与测评培训两个因素的分析表明, 就语言测评知识而言, 测评培训对此有提高作用, 而教学经验并不是显著影响因素^①, 这一情况可能与当下新手教师的专业学习和训练时间较近而在在职教师培训有所缺失有关。从调查对象参与测评培训的结果来看, 三年内参加过测评培训的人数在新手教师中占 51.28%, 在熟练型教师中占 46.34%, 在专家型教师中占 31.71%, 即使是在新手教师中, 测评培训的普及程度也仅在一半左右, 说明职前培训和在职培训中关于语言测评的学习均需要加强。

表 7 教学经验和参加测评培训因素对语言测评知识问卷得分的影响

语言测评知识问卷总分	新手教师		熟练型教师		专家型教师	
	未参加过测评培训	参加过测评培训	未参加过测评培训	参加过测评培训	未参加过测评培训	参加过测评培训
平均值	44.38	50.52	46.00	47.54	45.44	52.24
标准差	10.16	10.81	14.82	9.74	10.10	6.12
标准误	2.54	2.25	4.11	1.91	2.52	1.22

2. 语言测评实践

教师的语言测评实践中, 课程成绩一般由“课堂表现”、“课堂小测”、“课后练习”、“阶段测试”、“期末考试”组成。就课堂表现而言, 教师认为根据课堂互动情况对学生进行评价, 对考试成绩有一定预测性, 比较可靠。形成性评价在总成绩中比期中、期末考试等终结性评价所占比重更大, 教师同时认为期末学业考试是必要的, 规定无论学期中学生是否达到及格成绩, 都需要参加期末考试。

教师从事的语言测评相关工作主要包括“考试设计与命题”、“帮助学生备考”、“评分”、“交流反馈”。在命题方面, 与 Stiggins 的观点契合, 教师认为教学资源中已有的练习或测试题一定程度上减少了自主命题的需要^②。在备考方面, 学校开设 HSK 辅导课, 在综合课及各技能课上教师也会让学生练习 HSK 的题型, 反映了一定的高风险测试导向性^③。在评分和反馈方面, 教师关注到考试效度和分数意义的解释, 例如发挥考试的诊断作用, 为写作、课堂报告等主观试题制定评分标准等。

关于线上的语言测评实践, 教师使用多种线上会议平台、汉语学习以及数据收集平台进行教学与测评, 形式转变促使教师对于测评方式和效果进行思考和调整。传统的纸笔测试、课堂出勤率等评价方式需要完成模式的转变, 例如教师使用问卷星等工具收集测试结果, 在直播课上要求学生积极有效互动以确认课堂表现情况。作答方式需要适应线上测评, 例如高年级写作任务通过打字完成。评分标准发生一定变化, 例如学生的课堂报告在线上与线下两种模态的评分维度会发生变化。总之, 线上测试虽然在监督、实施等方面遇到问题, 但也有其优势, 例如学生学习数据的记录、测试批改和反馈等方面更加智能化, 发挥了计算机辅助学习的作用。为了保证线上考试公平有效, 面对学生地区时差的问题, 起初是通过调整在线考试的形式来解决, 后来通过完善编班和考试系统, 保证考生可以在同一时间作答, 为监考提供了可行条件。

实践背后是观念的指导, 根据访谈中教师表达的对于语言测评的观念, 绝大多数教师认同语言测评活动

^①Deborah Crusan, Lia Plakans, Atta Gebril, "Writing Assessment Literacy: Surveying Second Language Teachers' Knowledge, Beliefs, and Practices," 43-56.

^②Richard J. Stiggins, "Assessment Literacy," 534-539.

^③Tziona Levi, Ofra Inbar-Lourie, "Assessment Literacy or Language Assessment Literacy: Learning from the Teachers," 168-182.

的重要性,并持有积极态度,认为测评的功能是检验水平或学习成果,监督、规范学生学习,并针对性地改善教学。也有部分教师认为正式测试的目的是统计学生群体水平的量化信息,而在教师对学生的语言水平足够了解的情况下,测试成绩信息对教师来说是辅助性而不是必需的,甚至可能与教学关联较弱。此外,测评还与一些决策和风险相关,如分班、奖学金评审等。与 Brown 提出的四条测评观念假设^①对比,国际中文教师对语言测评的观念较为符合“测评对教学有促进作用”、“测评反映学生学习情况”,某些情况下“测评与教学无关”。

3. 语言测评反思

教师语言测评素养自评量表得分($M=47.43/79.38\%$, $SD=7.36$)和提升需求量表得分($M=48.06/80.1\%$, $SD=7.56$),均处于较高水平(各组量表评分的描述性数据见表 8、9)。教学经验对语言测评素养自评量表 $F(2, 116)=21.089$, $p=0.000$, $\eta^2=0.272$ 和提升需求量表 $F(2, 116)=7.227$, $p=0.001$, $\eta^2=0.113$ 结果的提升作用显著,与 Deluca 等结果相似^②;对于自评量表评分,三个教龄段之间均有显著差异;对于提升需求量表评分,教学经验的显著性影响体现在新手教师和专家型教师之间。测评培训因素对量表分数也有显著影响,自评量表 $F(1,117)=5.191$, $p=0.025$, $\eta^2=0.044$,提升需求量表 $F(1,117)=4.275$, $p=0.041$, $\eta^2=0.036$ 。教学经验和测评培训两个因素之间的交互作用不显著。虽然均有显著积极影响,教学经验对自评量表得分的效应量(0.272)明显大于测评培训因素(0.044),参加测评培训使教师的自我评价提升幅度较小,Deluca 等则发现未参加测评课程的教师自我评价比参加过的更高,认为这可能因为参加过测评培训的教师对相关概念复杂度的理解比未参加过的教师更深刻,因此对自身的评价趋于保守^③。

表 8 教学经验和参加测评培训因素对语言测评素养自评量表得分的影响

自评量表总分	未参加过测评培训			参加过测评培训		
	新手教师	熟练型教师	专家型教师	新手教师	熟练型教师	专家型教师
平均值	40.00	46.15	50.88	44.35	48.73	52.12
标准差	5.50	6.59	4.46	7.61	6.52	6.04
标准误	1.38	1.83	1.11	1.59	1.28	1.21

表 9 教学经验和参加测评培训因素对语言测评素养提升需求量表得分的影响

提升需求量表总分	未参加过测评培训			参加过测评培训		
	新手教师	熟练型教师	专家型教师	新手教师	熟练型教师	专家型教师
平均值	43.00	44.69	51.06	47.35	48.62	51.20
标准差	7.87	8.75	5.50	6.17	7.64	6.99
标准误	1.97	2.43	1.37	1.29	1.50	1.40

在教师对语言测评素养发展的看法中,教师认为教学经验对于课堂测评中把握学生语言水平、学习重难点方面的能力提升有较大作用,有利于设计测评任务和命题,而与更加专业的测量理论和量化分析技能的提高不直接相关,后者需要通过专门的学习和运用来实现,在测试方面较为专业的知识和技能通常是教学研究所用。参加与测试相关的工作为教师提供了思考和学习的契机,例如教师认为为学生提供 HSK 备考的辅导使其对语言测试有了更多的了解。

①Gavin T. L. Brown, "Teachers' Conceptions of Assessment: Implications for Policy and Professional Development," 301-318.

②Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe- McEwan, Ulemu Luhanga, "Teachers' Approaches to Classroom Assessment: A Large-scale Survey," 355-375.

③Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe- McEwan, Ulemu Luhanga, "Teachers' Approaches to Classroom Assessment: A Large-scale Survey," 355-375.

教师认为自身的语言测评素养有待提升,且接受过测评相关培训的教师在教学工作中会发现其他教师缺乏语言测评素养的一些做法。与已有研究相似^①,国际中文教师对语言测评素养的提升需要具体可行的实践指导,包括测试原则(标准参照、常模参照测试等)、测试命题指导(如何利用大纲、如何提升效度和信度等)、提供学习建议(能力、知识诊断等)等方面。

(二)讨论

基于以上研究结果的分析,本研究主要的发现和解释如下。

1.国际中文教师对语言测评知识的掌握程度为中等(62.63%),对客观知识的掌握并不充足,这与已有研究结果相似^②,语言测评素养自评量表和提升需求量表的结果较高(79.38%,80.1%),说明国际中文教师对语言测评素养的评价可能存在“不实际的乐观态度”^③。

2.从影响因素来看,参加测评培训对于语言测评知识得分有积极影响,说明现有的测评培训对于普及语言测评知识有一定效果,主要体现在环境(指导性标准和大纲等)和实践(测评活动中的具体做法)方面。Mertler使用“课堂测评素养问卷(CALI)”测量教师测评素养,发现课堂教学经验对于提升语言测评素养在多个方面都有显著作用^④,而本研究发现教学经验对语言测评知识得分基本没有产生显著影响,但教学经验对自评和提升需求量表的结果有显著影响,结果反映出语言测评“经验丰富而非专家”^⑤的现象。从访谈结果看,语言测评素养对于国际中文教师是一个尚未普及的概念,诚然,教学经验可以使教师更加准确地诊断学生能力,但在没有专业知识的指导下依然存在很多教师测评理念并不合理的情况,不能正确处理测试公平性、有效性的问题。

3.国际中文教师的语言测评实践注重形成性评价,也存在一定的高风险测试导向。教师对语言测评活动的看法整体是积极的,也存在认为教学经验可以代替正式测试的观点,对于教师来说,测评依然是国际中文教学中的边缘成分^⑥,更多与教学结合而不具有独立的专业性。部分教师认为正式的语言测试与测评专业知识的重要性不强,此观点可以解释问卷结果中语言测评知识方面教学经验不是显著影响因素(因为教师在教学过程中没有足够的动力学习这方面知识),也反映出对于国际中文教师来说,促进教学的评价与标准化测试两方面的能力是相对独立的^⑦,前者更加受到教师重视。

4.国际中文教师对语言测评素养的提升持积极态度,认为有必要提升测评培训质量,更新在职学习内容。教学经验丰富的教师对语言测评素养的自我评价更高,对提升需求的评分也较高,参加测评培训也使提升需求更高。与DeLuca等使用“课堂测评方法调查问卷(ACAI)”调查结果相比^⑧,教学经验因素在提升需求量表上表现出更强的促进作用,强调了持续性、与工作结合的测评学习的重要性。教师认为对语言测评素养的提升可以从职前培训、实践和研究指导等方面有针对性地帮助提高教师语言测评素养和相关研究兴趣。教师语言测评素养的提升需要政策环境支持,对于语言测评素养在实践中的巩固和提升,以及创新性测评模式的运用,学校或机构需要提供支持条件,以减少如年轻教师没有话语权、测评上的创新做法受到管理条件

①Margaret E. Malone, “The Essentials of Assessment Literacy: Contrasts between Testers and Users,” 329-344.

②郑东辉《中小学教师评价素养状况:来自Z省的报告》,《全球教育展望》2010年第2期,第31—36、42页; Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe-McEwan, Ulemu Luhanga, “Teachers’ Approaches to Classroom Assessment: A Large-scale Survey,” 355-375; 江进林《高校英语教师测评素养现状及其影响因素研究》,《外语界》2019年第6期,第18—26页。

③Christopher DeLuca, Don A. Klinger, “Assessment Literacy Development: Identifying Gaps in Teacher Candidates’ Learning,” *Assessment in Education* 17, no. 4 (October 2010): 419-438.

④Craig A. Mertler, “Secondary Teachers’ Assessment Literacy: Does Classroom Experience Make a Difference?” 76-92.

⑤Deborah Crusan, Lia Plakans, Atta Gebril, “Writing Assessment Literacy: Surveying Second Language Teachers’ Knowledge, Beliefs, and Practices,” 43-56.

⑥Esmat Babaii, Fatemeh Asadnia, “A Long Walk to Language Assessment Literacy: EFL Teachers’ Reflection on Language Assessment Research and Practice,” 745-760.

⑦Angela Scarino, “Language Assessment Literacy as Self-awareness: Understanding the Role of Interpretation in Assessment and in Teacher Learning,” 309-327.

⑧Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe-McEwan, Ulemu Luhanga, “Teachers’ Approaches to Classroom Assessment: A Large-scale Survey,” 355-375.

制约等问题^①。

五 总结与建议

本研究反映了国际中文教师多维度、多层面的语言测评素养,研究结果可以为国际中文教师的师资培养和教师专业发展提供参考。研究使用了选择题和量表两种问卷方式调查语言测评素养,二者反映的结果可以对比和互补。“语言测评知识问卷”内容以点状呈现,通过知识的样本来反映教师对语言测试概念和原则的理解以及特定情境下的选择;而“语言测评素养评价量表”的内容以能力描述来呈现,让教师评价自身所具备的语言测评相关能力和提升需求。两份问卷的结合可以从不同视角解释信息,弥补单一调查工具的不足^②。本研究的访谈包含教师对日常语言测评活动和相关工作的描述,也包含教师对语言测评素养概念的理解和现状反思。整体来说,本研究的开展遵循认识到反思的顺序,研究框架体现了 Xu & Brown 实践中的教师测评素养的发展循环^③。

语言测评素养是理论知识、实践能力和价值态度等多方面融合的表现,对于素养的水平不能仅以“有无”描述,我们从知识、实践和反思三方面来尝试解读国际中文教师的语言测评素养。知识方面,国际中文教师的专业学习、职前培训等经历保证了一定的知识基础,但是问卷与访谈结果显示,具有系统的语言测评知识体系的教师较少,且教师对相关概念(信度、效度等)的理解比较浅显,难以对实践有指导作用。在实践方面,教学经验的作用得以突显,有经验的教师比较重视以评促学的理念,对学生实施多元化、追踪性的评价。然而语言测评实践的情况随教师个人、教学环境变化较大,很多新手教师由于工作内容有限,设计整体测评方案和实施测评活动的需求较少。而且,无论对于新手教师还是经验教师,在测评上花费精力有时被视为负担。在反思方面,教师大多认同测评对于教学的影响力以及语言测评素养的重要性,且指出参与本研究也是一个反思语言测评的机会,将语言测评的实践纳入教学反思中更有利于教师更有意识地调整和完善语言测评方面的工作。

针对语言测评素养在这三个方面的表现,我们提出以下提升模式与建议。(1)对语言测评培训课程的目标进行重新定位。根据教师语言测评素养的特点以及需求分析,改进已有的测评课程,在新手教师的培养阶段,要打好测评原理和原则的理论知识基础,并提供实践指导。要为在职教师提供适应具体测评环境和实践需求的学习机会。例如,课堂测评是教师语言测评工作的重要阵地,在测评培训中可以加入更多前沿的课堂测评方式及其成绩报告和反馈的应用等。又如在线上教学资源丰富的情况下,教师需要懂得如何评估、选择和整合现有的测试和练习资源。(2)将测评融入教研工作中,通过学徒、合作等方式在实践中发展测评素养。强调测评工作的规范和科学性,如通过教师合作命题等方式调动其思考和学习测评方式的积极性。对于高校在职中文教师,可以鼓励其将教学与科研相结合学习语言测评相关知识,并应用一些创新的测评方式促进教学。

本研究尚存在不足之处。在研究对象上,样本数量可以更大,对于教师的教学经验和培训经历分组可以更加清晰合理。在研究方法上,本研究问卷所包含的问题和答案并没有在真实教学情况进行检验,因此问卷虽一定程度上可以反映教师在语言测评专业方面的知识和素养情况,但与具体教学环境的契合程度有待进一步考察。访谈结果基于教师的自我报告,可能与具体实践情况有偏差,因此可以采取观察课堂或测评培训课程中教师的实际操作的方式来对语言测评素养的表现有更深入的理解。

今后有关国际中文教师语言测评素养的研究可以从以下几方面来进行拓展:(1)通过课堂观察、案例研究等方式更加深入地了解教师在具体实践中的语言测评素养;(2)结合现状与教师反映的需求,设计开展相应的培训项目,并据此探索与验证语言测评素养的发展方式;(3)在整合语言测评素养内涵的基础上,制定国

^①Ricky Lam, "Teacher Assessment Literacy: Surveying Knowledge, Conceptions and Practices of Classroom-based Writing Assessment in Hong Kong," 78-89.

^②Chad M. Gotch, Brian F. French, "A Systematic Review of Assessment Literacy Measures," *Educational Measurement: Issues and Practice* 33, no. 2 (June 2014): 14-18; Christopher DeLuca, Adelina Valiquette, Andrew Coombs, Danielle LaPointe-McEwan, Ulemu Luhanga, "Teachers' Approaches to Classroom Assessment: A Large-scale Survey," 355-375.

^③Yueting Xu, Gavin T. L. Brown, "Teacher Assessment Literacy in Practice: A Reconceptualization," 149-162.

际中文教师语言测评素养的测量工具并进行有效性验证和完善。

Investigating the Language Assessment Literacy of College Teachers of Chinese as a Second Language: Knowledge, Practice, and Reflection of Language Assessment

Zhang Yanli¹, Wang Xiaozhu²

1. School of Chinese Studies and Exchange, Shanghai International Studies University, Shanghai 200083, China

2. Institute of Educational Policy and Evaluation of International Students, Beijing Language and Culture University, Beijing 100083, China

Abstract: Language assessment literacy (LAL) is an essential professional competence for language teachers to effectively design, interpret, and implement language assessments. This study examines the LAL of Chinese as a Second Language (CSL) college teachers by exploring their knowledge, practice, and reflection regarding language assessment. The findings reveal that CSL teachers possess an intermediate level of language assessment knowledge, displaying a grasp of fundamental concepts and principles, but also exhibiting knowledge gaps and misunderstandings. The study shows that participation in assessment training programs significantly enhances language assessment knowledge. In practice, CSL teachers recognize the importance of formative assessment, yet also demonstrate a tendency to align with high-stakes tests. Both teaching experience and assessment training positively influence teachers' perceived level of LAL and learning need. Teachers express a positive attitude towards improving their LAL, emphasizing the need for practical guidance offered by language assessment training.

Key words: language assessment literacy; teachers of Chinese as a Second Language; language assessment knowledge; teacher's professional development

[责任编辑:唐 普]