

● 外语教学: 能力量表研发专题

## 国外批判性思维测评研究进展\*

马利红 刘坚

(天津职业技术师范大学, 天津 300222; 北京师范大学, 北京 100875)

**提 要:** 作为 21 世纪技能之一, 批判性思维在外语教育界越来越受到关注, 但是批判性思维教学只有和批判性思维评价相结合才能真正培养学生的批判性思维。本文从测评内容、测评形式、测评题型、信度和效度证据等方面综述国外批判性思维测评研究的新进展, 旨在为我国外语界开展批判性思维评价研究提供参考。

**关键词:** 批判性思维测评; 效度; 信度; 测评形式; 外语教育

中图分类号: H319

文献标识码: A

文章编号: 1000-0100(2019)03-0064-5

DOI 编码: 10.16263/j.cnki.23-1071/h.2019.03.011

### The Development of Overseas Critical Thinking Assessment

Ma Li-hong Liu Jian

(Tianjin University of Technology and Education, Tianjin 300222, China; Beijing Normal University, Beijing 100875, China)

As one of the 21<sup>st</sup> century skills, critical thinking has received more and more attention in foreign language education. However, critical thinking teaching can cultivate students' critical thinking only by being combined with critical thinking assessment. This paper reviewed the new development of overseas critical thinking assessment based on assessment content, evaluation form, assessment items, the reliability and validity evidence, etc. It is hoped that the paper will provide insight into critical thinking assessment research in foreign language education in China.

**Key words:** critical thinking assessment; validity; reliability; evaluation form; foreign language education

#### 1 引言

外语界近几年与批判性思维有关的研究数量不断上升, 且主要集中在批判性思维培养与外语技能的融合方面 (Razaei et al. 2011: 769-777)。但是批判性思维教学只有和科学的批判性思维评价相结合才能真正培养学生的批判性思维 (Abra-mi et al. 2014: 275-314)。作为提升批判性思维培养质量的有效方式, 批判性思维测评研究成为外语教育领域比较重要的研究课题。

目前, 我国外语界对于批判性思维的测评研究以外语类大学生思辨能力量具 (文秋芳等 2010: 19-26)、英语专业硕士生思辨倾向量具

(张莎 文秋芳 2017: 110-114) 为代表, 对其他学段外语学习者批判性思维的测评研究尚不足。本文探讨国外批判性思维测评研究的新进展, 旨在为我国外语界未来相关研究指明发展方向。

#### 2 测评内容

从测评内容上, 国外批判性思维测评分为批判性思维人格倾向测评、批判性思维技能测评和综合测评。

##### 2.1 批判性思维人格倾向测评

批判性思维人格倾向是一个人进行批判性思维的态度、特质和倾向 (Bensley et al. 2016: 158-

\* 本文系全国教育科学“十三五”规划教育部重点课题“基础教育阶段英语学科素养的内涵、培养路径和测评指标研究”(DHA160310)和2019年教育部人文社会科学规划基金项目“外语写作中的批判性思维评价效度研究”(19YJA740038)的阶段性成果。

作者电子邮箱: professorliu9506@126.com(刘坚)

168)。有学者认为批判性思维人格倾向常常包括寻求真理、开放性、系统性、自信心、好奇心、认知成熟度等(Facione, Facione 1992)。

常见的批判性思维人格倾向测评工具包括加利福尼亚心理动机测评(CM3)和加利福尼亚思辨人格倾向量表(CCTDI)等。CM3是针对美国中小学生的批判性思维人格倾向量表,包括学习导向(Learning Orientation)、创造性问题解决(Creative Problem Solving)、精神集中(Mental Focus)、认知完整性(Cognitive Integrity)4个维度,研究表明这4个维度与学生的学习动机和学业成绩显著正相关(Giancarlo et al. 2004: 347-364)。CM3测量的批判性思维人格倾向不局限于特定学科课程,而是测量学生智力活动(如推理)的认知参与度和内在动机。CCTDI是应用广泛的批判性思维人格倾向测评工具,主要应用于大学生和中学高年级学生(Facione, Facione 1992)。

近些年来,国际文献中报告研究者对批判性思维人格倾向量表做进一步开发和验证,如批判性思维人格倾向量表(CTDS)的发展和心理测量评价过程。CTDS的题目由批判性思维人格倾向工具改编,运用CTDS分别对两组被试(一年级本科生和研究生)进行心理测量评价,在第一个研究中,探索性因子分析结果表明CTDS是两因子模型:批判性开放和反思性怀疑;第二个研究中,多组验证性因素分析进一步验证两因子模型,而且多组验证性因子分析结果表明两组学生理解题目的路径相似(Sosu 2013: 107-119)。教育学家、心理学家都认为CTDS是测评批判性思维人格倾向的有效工具(同上)。

### 2.2 批判性思维技能测评

对批判性思维技能及子技能比较有影响的分类是美国哲学协会德尔菲项目提出的解释、分析、评价、推理、阐释、自我调节6项技能和16项子技能(Facione, Facione 1992)。基于批判性思维是“一种认知,有助于决定做什么或相信什么”,研究者开发康奈尔批判性思维测评(CCTT),题目涉及假设、可信度、演绎、归纳、推理5个方面,CCTT将测试题目融入真实生活场景中,而且整套题目的测试场景是连贯的(Ennis, Millman, Tomko 1985)。根据元分析结果,CCTT是应用最广泛的批判性思维测评工具(Abrami et al. 2014: 275-314)。

根据德尔菲项目对批判性思维测评的建议,研究者开发加利福尼亚思辨技能测评工具(CCTST)。之后,研究者根据德尔菲报告的测评指南进一步发展CCTST,旨在测评批判性思维的

过程性技能,如演绎、评价、推理和论证等。

### 2.3 批判性思维综合测评

批判性思维是多维构念,既包括批判性思维技能,也包括批判性思维人格倾向和元认知(Bensley et al. 2016: 158-168),因此批判性思维综合测评是一项艰巨任务。哈尔彭批判性思维测试(Halpern Critical Thinking Assessment, HCTA)是基于日常生活场景将批判性思维人格倾向和技能融合在一起的综合测评,大量样本已证明HCTA具有较好的信度和效度(Halpern 1998: 449-455)。最近开发的HCTAES(Halpern Critical Thinking Assessment Using Everyday Situations)从5个维度评价大学生的批判性思维:推理能力、论证分析能力、假设检验能力、应用能力、决策和解决问题的能力,题目涉及25个日常生活情境,每个情境下包含一道选择题和一道开放题,进一步完善HCTA(同上2006)。

## 3 测评形式

从测评形式上,国外批判性思维测评包括嵌套于特定学科的批判性思维测评和通用型批判性思维测评。

### 3.1 特定学科批判性思维测评

嵌套于特定学科的批判性思维测评用于评价某一特定学科领域的批判性思维,如外语界学者根据学生的外语写作表现评价学生的批判性思维(Stapleton 2002: 250-257)、批判性思维特质调查CTAS(Critical Thinking Attribute Survey)主要测试与教育和教师发展相关的特定学科批判性思维技能(Forawi 2016: 52-62)。语篇测试(essay test)和“真实情景评价”被认为是在特定学科领域提高批判性思维评价效度的两种测评方法(Tynja et al. 1998: 173-189)。

目前,嵌套于特定学科的批判性思维测评已经在下列学科领域得到发展,如心理学(Bensley et al. 2016: 158-168)、生物学(McMurray et al. 1991: 183-192)、外语教育(Stapleton 2002: 250-257)等。但是目前这些嵌套于特定学科的批判性思维测评主要用于预测学科成绩,很少真正用于评价学习者的批判性思维技能和倾向。

### 3.2 通用型批判性思维测评

通用型批判性思维测评用于评价学生的通用批判性思维技能,如CCTT和CCTST。剑桥思维能力测评TSA(Cambridge Thinking Skills Assessment)也是一项评价学生学习任何专业都需要的通用思维技能的测评工具,目前已被剑桥大学、牛

津大学和伦敦大学学院采用。

#### 4 测试题型

国外批判性思维测试采用选择题、开放题、选择题+开放题、语篇测试、表现性评价等形式。

##### 4.1 选择题

选择题具有内容覆盖面广、评分标准客观、评分速度快等独特优势,可为评价批判性思维的基本方面提供证据,因此有人建议在开发批判性思维测试题时使用选择题。但研究证明利用选择题测评批判性思维受被试批判性思维复杂程度和测评理念影响,只能测量认知水平,难以测试批判性思维的核心内容,难以评价被试对解决复杂问题的态度(Ku 2009: 70-76, Norris 1989: 21-26)。

根据评价批判性思维测评工具的3个标准:一是测评是否基于批判性思维的科学概念,二是测评内容是否全面,三是测评题目的难度是否符合学生水平,研究者发现目前很多批判性思维测评缺乏综合性内容,特别是选择题,往往漏掉批判性思维中的重要信息(Ennis 1993: 179-186)。

尽管选择题测评批判性思维有诸多弊端,但综合科学性、有效性、评分一致性、经济性和可行性等多方面考虑,改善选择题仍具有重要的实践意义(Norris 1989)。因此,很多批判性思维测试仍使用选择题形式,比如美国大学学业水平测试CAAP(Collegiate Assessment of Academic Proficiency test)就采用选择题评价学生对特定语篇进行分析、评价和归类的能力。2015年11月ETS推出的评估大学生学习成果的最新批判性思维测试也采用选择题形式,在每段语料后给出几个问题,主要考查分析和综合能力,分析能力包括评估证据本身及其使用、分析和评估论证过程、理解论证语言、区分有效论证和无效论证,综合能力包括展开有效论证等(刘欧 2016: 13-17)。

##### 4.2 开放题

采用开放题测评批判性思维越来越引起学界关注(马利红 2018)。例如,作为一种开放性批判性思维测评工具,EWCTET(The Ennis-Weir Critical Thinking Essay Test)适用于高中及大学生,评价被试辨别语篇中的推理漏洞以及如何捍卫自己的立场,即测评被试的分析能力和在真实情景中的反应能力和论证能力(Ennis, Weir 1985)。

影响力较大的开放性批判性思维测评方法是3种基于语篇的批判性思维测评方法(Ennis 1993: 179-186)。一是高复杂结构(high structure),运用议论文测评批判性思维,即提供一篇

论证性文章,标出段落,但大部分段落中有论证错误,要求学生每一段及整篇文章中隐含的写作者思维方式,进行评价,并论证自己为何这样评价。EWCTET就是高复杂结构语篇测验(high structure essay test)的例子。二是中等复杂结构(medium structure),也是采用议论文测评批判性思维,即提供一篇论证性文章但文章结构相对简单,要求学生文章主题进行论证或者辩护,但是不具体阐述为何这样论证或辩护,可采用综合评分或分析评分方法。目前,国外有些大学的高级定位考试(College Board AP tests)一般采用这种题型。三是微复杂结构(minimal structure),用于测评的文章结构较简单,要求考生只回答一个问题或阐述一件事,比如要求学生就某个感兴趣的话题或问题发表看法或捍卫自己的立场,这种题型可以给学生更多发挥空间,但是为教师提供的诊断性信息较少,也可采用综合评分和分项评分方式。美国的伊利诺斯批判性思维作文大赛(The Illinois Critical Thinking Essay Contest)使用这种测试题型。尽管批判性思维语篇测试这种开放性测评模式有利于展示被试思维过程中的认知能力和认知倾向,但是很多人也认为特定情境和严谨结构限制了被试作答,难以充分显示被试的思维倾向,而且主观评分深受评分员信度影响(Bridgeman et al. 2012: 27-40)。

为测量批判性思维的教学成效,学者们积极探索更多开放性测评方式,如档案袋评价、采访、小组调查、假设情景应答、学生自我报告、认知测试等(Halpern 2006)。另外,研究者还建议使用半结构性开放题,即提供真实生活场景,让学生批判性表达自己的思想(Ennis 1993: 179-186),这种测评题型可以考查批判性思维的更多方面。“开放题可反映被试的思维过程,可考查被试思维过程的清晰性、相关性与逻辑性,还可考查辩证思维的深刻性和灵活性,但是开放题也有一定的局限性,比如评分带有一定主观性,评分速度慢,信度低等”(文秋芳 2012)。

##### 4.3 选择题与开放题相结合

选择题和开放题在测评批判性思维时各有局限性,比如选择题在命题时费时费力,难以全面、有效地测试出批判性思维技能和倾向,开放题则在评分时耗费精力,信度难以保障,开放题中的测试情景又太具体、主观性偏强(Liu et al. 2014: 1-23)。于是,有研究者提出一个折中办法,即采用选择+构造反应(constructed response)的测评模式,让学生在选出答案后写出原因,由于命题者和被试的立

场与观念不同,被试可以论证自己的答案(Ennis 1993: 179-186)。鉴于选择+构答反应这种题型综合性强,可弥补选择题的不足,允许学生背景差异和对题目阐释的差异存在,很多研究者力荐采用选择+构答反应的形式命制批判性思维题目。研究表明尽管构答反应题型可为批判性思维测评提供真实情景,但是表面效度很强,作答时间等同的情况下,构答反应题的信度低于选择题(Lee et al. 2011: 115-136)。

Halpern研发的HCTA和HCTAES将选择题和构答反应相结合,每个情境下有一道选择题和一道开放题,测量被试的不同认知能力,选择题主要考查批判性思维技能,开放题主要考查批判性思维倾向。另外,EPP(ETS Proficiency Profiles),CLA+都是采用选择题和构答反应相结合的形式(Educational Testing Service 2010)。Ku(2009: 70-76)指出选择题、调查问卷可能不是最有效的测评批判性思维的方式,并提出包括多项选择和开放题的综合性测评方式以监测被试选择某一选项时的思维过程,与Ennis(1993: 179-186)和Halpern(2006)的测评理念一致。

#### 4.4 表现性评价

近些年来,国际上出现通过完成某项任务测评批判性思维能力的方法,即表现性评价。自然观察是常见的表现性评价方法之一,如受过培训的观察者以一个人或一组人的活动为中心,进行观察并记录,描述一系列事情。结构稍复杂的表现性评价是运用学生档案袋记录学生批判性思维的日常发展和进步情况。尽管表现性评价是测评批判性思维的好办法,但是这类评价的效度难以建立(Ennis 1993: 179-186)。

目前国际上已有的结构较严谨的表现性评价是美国国家教育进展评估(NEAP)开发的拓展性评价。2002年,美国教育援助委员会通过表现性评价测评大学毕业生的批判性思维技能,包括选择题和开放性写作题,其中开放性写作题比较成熟,要求学生阅读分析相关语篇后做出批判性回应。现在通过表现性评价测评学生的批判性思维已被拓展应用到美国中学生群体。

#### 5 信度和效度

批判性思维具有多维特征,但是现有批判性思维测评大多报告分维度的得分表现。虽然分维度得分可提供批判性思维某一技能的详细信息,但这些维度的信度并不高,维度划分并没有足够的实证证据(Liu et al. 2014: 1-23)。更有甚者,

基于60项研究的元分析表明批判性思维其实只有一个单因子(Bernard et al. 2008: 15-22)。

批判性思维测评的效度研究主要通过与其他认知技能的相关性来验证,比如批判性思维与一般认知技能(如SAT和GRE)具有中度相关,与GPA和课程成绩中度相关(Halpern 2006),而且具有较高批判性思维的人比具有较低批判性思维的人生活更加积极。关于效度验证的研究表明批判性思维测评研究的数量和质量差异较大,批判性思维测评的普遍问题包括维度划分证据不充分,子维度信度低,效度验证证据不充分等(Liu et al. 2014: 1-23)。

#### 6 结论与启示

批判性思维测试工具和测评方式的发展说明批判性思维测评越来越多样化、学科化和动态化,为我们开发研制适合中国外语学习者的本土化测评工具提供很好的借鉴。未来批判性思维测评要进行情景化、个性化和多样化的探索,量化测评和质性评价相结合,以更加详实具体的数据展示学生的批判性思维技能和人格倾向。

思维除具有普遍性特征外,还受文化因素影响(文秋芳 2012)。国内学者对批判性思维的测评研究缺乏对文化因素的探讨,难以和国际上同类研究进行比较。因此,研究我国外语学习者的批判性思维能力,不能一味照搬西方理论,测评工具也不能停留在对国外工具的翻译和修订上,应结合我国学生的思维特点,开发适合我国文化背景和学段特征的批判性思维测评工具。

#### 参考文献

- 刘欧. 高校学生学习成果测评的历史、现状和前瞻[J]. 中国考试, 2016(11). || Liu, O. Student Learning Outcomes Assessment in Higher Education: A Historical Review, Current State, and Future Directions[J]. *China Examinations*, 2016(11).
- 马利红. 国外批判性思维开放题测评的发展及启示[J]. 中国考试, 2018(3). || Ma, L.-H. The Development of Open Questions in Overseas Critical Thinking Assessment and Its Implications[J]. *China Examinations*, 2018(3).
- 文秋芳. 中国外语类大学生思辨能力现状研究[M]. 北京: 外语教学与研究出版社, 2012. || Wen, Q.-F. *A Study on the Current Situation of Chinese Foreign Language Students' Critical Thinking Ability* [M]. Beijing: Foreign Language Teaching and Research Press, 2012.
- 文秋芳 王海妹 王建卿 赵彩然 刘艳萍. 我国外语

- 类大学生思辨能力量具的修订与信效度检验研究[J]. 外语界, 2010(4). || Wen, Q.-F., Wang, H.-M., Wang, J.-Q., Zhao, C.-R., Liu, Y.-P. Revision and Reliability and Validity Test of Chinese Foreign Language Students' Critical Thinking Ability Measurement [J]. *Foreign Language World*, 2010(4).
- 张莎 文秋芳. 英语专业硕士生思辨倾向的量具构建与实证调查[J]. 外语学刊, 2017(2). || Zhang, S., Wen, Q.-F. Inventory Development and Investigation of MA English Majors' CTDs [J]. *Foreign Language Research*, 2017(2).
- Abrami, P., Bernard, E., Borokhovski, D., Waddington, C., Wade, C., Persson, T. Strategies for Teaching Students to Think Critically: A Meta-analysis [J]. *Review of Educational Research*, 2014(2).
- Bensley, D. A., Rainey, C., Murtagh, M. P. Closing the Assessment Loop on Critical Thinking: The Challenge of Multidimensional Testing and Low Test-taking Motivation [J]. *Thinking Skills and Creativity*, 2016(21).
- Bernard, R., Zhang, D., Abrami, P., Sicoly, F., Borokhovski, E., Surkes, M. Exploring the Structure of the Watson — Glaser Critical Thinking Appraisal: One Scale or Many Subscales? [J]. *Thinking Skills and Creativity*, 2008(3).
- Bridgeman, B., Trapani, C., Attali, Y. Comparison of Human and Machine Scoring of Essays: Differences by Gender, Ethnicity, and Country [J]. *Applied Measurement in Education*, 2012(1).
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., Almeida, L. S. The Halpern Critical Thinking Assessment and Real-world Outcomes: Cross-national Applications [J]. *Thinking Skills and Creativity*, 2012(7).
- Educational Testing Service. *ETS Proficiency Profile User's Guide* [M]. Princeton: Author, 2010.
- Ennis, R. H. Critical Thinking Assessment [J]. *Theory into Practice*, 1993(3).
- Ennis, R. H., Millman, J., Tomko, T. N. *Cornell Critical Thinking Tests* [M]. Seaside: Critical Thinking Company, 1985.
- Ennis, R. H., Weir, E. The Ennis-Weir CT Essay Test: An Instrument for Testing and Teaching [R]. Pacific Grove: Midwest Publications, 1985.
- Facione, P. A., Facione, N. C. *The California Critical Thinking Disposition Inventory* [M]. Millbrae: California Academic Press, 1992.
- Forawi, S. A. Standard-based Science Education and Critical Thinking [J]. *Thinking Skills and Creativity*, 2016(20).
- Giancarlo, C. A., Blohm, S. W., Urdan, T. Assessing Secondary Students' Disposition Toward Critical Thinking: Development of the California Measures of Mental Motivation [J]. *Educational and Psychological Measurement*, 2004(2).
- Halpern, D. F. Teaching Critical Thinking for Transfer Across Domains: Dispositions, Skills, Structure Training, and Metacognitive Monitoring [J]. *American Psychologist*, 1998(4).
- Halpern, D. F. Halpern Critical Thinking Assessment Using Everyday Situations: Background and Scoring Standards [R]. Claremont: Claremont McKenna College Press, 2006.
- Ku, K.-Y. Assessing Students' Critical Thinking Performance: Urging for Measurements Using Multi-response Format [J]. *Thinking Skills and Creativity*, 2009(1).
- Lee, H. S., Liu, O.-L., Linn, M. C. Validating Measurement of Knowledge Integration in Science Using Multiple-choice and Explanation Items [J]. *Applied Measurement in Education*, 2011(2).
- Liu, O.-L., Frankel, L., Katrina, C. R. Assessing Critical Thinking in Higher Education: Current State and Directions for Next-Generation Assessment [J]. *ETS Research Report Series*, 2014.
- Norris, S. P. Can We Test Validly for Critical Thinking? [J]. *Education Researcher*, 1989(9).
- Sosu, E. M. The Development and Psychometric Validation of a Critical Thinking Disposition Scale [J]. *Thinking Skills and Creativity*, 2013(9).
- Stapleton, P. Critical Thinking in Japanese L2 Writing: Rethinking Tired Constructs [J]. *ELT Journal*, 2002(3).
- Tynja-la, P. Traditional Studying for Examination Versus Constructivist Learning Tasks: Do Learning Outcomes Differ? [J]. *Studies in Higher Education*, 1998(21).

定稿日期: 2019-04-10

【责任编辑 陈庆斌】