



国际中文教材动态评价模型构建研究

梁 宇

摘要:构建国际中文教材动态评价模型、实现网络实时诊断性评价是国际中文教育标准体系建设、提升我国国际传播力影响力的重要举措。本文立足第五代评价观和经典教材观,确立了中外融通、素养提升、教学适用、应用实践、科学规范、技术赋能六方面的评价理念,引进机器评价、诊断评价、循证评价等先进评价手段,综合运用多元研究方法,构建了国际中文教材动态评价模型。该模型分为规范评价、定量评价、定性评价、附加评价四部分,包括价值取向、教学适用、内容编排、外观配置 4 个一级指标、20 个二级指标和 60 个三级指标。在评价内容方面,该模型的特点包括坚守中华文化立场,打造柔性话语表达,注重两种能力培养,聚焦中文二语教学,突出教学适用评价,深化教育服务理念。在评价方式方面,该模型引入链式思维,强调动态开放,重视诊断反馈,探索综合性、过程性和发展性评价模式。

关键词:国际中文教育;国际中文教材;教材评价;动态评价;评价模型

DOI: 10.13734/j.cnki.1000-5315.2023.0619

收稿日期:2023-03-21

基金项目:本文系教育部人文社会科学重点研究基地重大项目“国际中文教育数字资源综合评价理论与方法研究”(22JJD740016)、国家语委重点项目“‘中文+职业技能’教学资源建设研究”(ZDI145-35)、教育部语合中心国际中文教育研究重大课题“国际中文教材评价指标体系构建”(21YHZD05)的阶段性研究成果。

作者简介:梁宇,女,天津人,语言学及应用语言学博士,北京语言大学国际中文教育研究院编审、博士生导师,研究方向为国际中文教育、国际中文教学资源研究,E-mail: liangyu@bleu.edu.cn。

一 引言

国际中文教材是推动国际中文教育高质量发展的关键要素,是传播中国语言文化的重要载体。通过教材评价进一步提升国际中文教材质量,对我国建设高质量国际中文教育体系、增强中华文明传播力影响力具有重要意义。2018 年全国教育大会以来,中共中央、国务院进一步加强教材建设和管理,明确教材建设是国家事权。2020 年教育部印发的《全国大中小学教材建设规划(2019—2022 年)》明确提出,“构建教材建设质量保障体系”^①。同年,中共中央、国务院印发的《深化新时代教育评价改革总体方案》提出,“完善教材质量监控和评价机制”^②。2021 年公布的《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》指出,要构建“国际中文教育标准体系”^③。党的二十大报告进一步强调,要“加快建设高质量教育体系”,“推动中华文化更好走向世界”^④。可见,构建国际中文教材动态评价模型(以下简称“评价模型”)是国际中文教育标准体系建设服务党和国家教育方针、提升我国国际传播能力的重要内容。

国内外二语教材评价研究主要包括两方面内容。一是评价理论研究。关于教材评价的本质主要有四种观

①《描绘新时代教材建设蓝图〈全国大中小学教材建设规划(2019—2022 年)〉发布》,《中国教师报》2020 年 1 月 15 日,第 1 版。

②《中共中央 国务院印发〈深化新时代教育评价改革总体方案〉》,《中华人民共和国国务院公报》2020 年第 30 号,第 13 页。

③《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》,《人民日报》2021 年 3 月 13 日,第 10 版。

④习近平《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》,人民出版社 2022 年版,第 34、46 页。

点:“对照观”认为,“教材评价是将教材使用主体需求分析与教材分析对照的过程”^①;“选择观”提出,“教材评价就是要帮助教师在各种教学环境下为各种水平的学生挑选最合适的教材”^②;“效果观”指出,“教材评价取决于教材的课堂使用效果”^③;“判断观”则强调,“教材评价是对教材价值的判断”^④。此外,学界对二语教材评价的原则、类型、主体、方法等均有较为全面的阐述。二是评价量表研制。前期研究一方面构建了丰富的教材评价量表,1970—2008年国外英语教材评价量表多达48个^⑤,2011—2021年改编或新开发的英语教材评价量表又有36个^⑥。我国学者研制的二语教材评价量表也有一定代表性^⑦。另一方面,评价量表的研制方法多元,早期的研制方法以理论建构为主,进入21世纪,评价量表更加注重科学性,研制方法主要有文本分析、问卷调查、焦点团体访谈等。总体来看,前期研究对二语教材评价理论和量表研制进行了有益探索,奠定了良好基础,但教材评价始终无法突破静态的量表式评价,评价体系的研制方法较为单一,评价手段较为传统。

本研究立足第五代评价观和经典教材观,从确立评价理念入手,综合运用多元研究方法构建评价模型,基于教材大数据,引入机器评价、诊断评价、循证评价等先进评价手段,探索综合评价和过程评价的路径,以期实现动态化、个性化和多元化的教材评价,为提升教材编写质量提供科学依据。

二 理论基础

(一)评价观

现代评价理论经历了四代演进,它们是以工具导向、测量为标志的第一代评价,以目标导向、描述为特征的第二代评价,以决策导向、判断为特征的第三代评价,以价值导向、建构为特征的第四代评价^⑧。之后,西方学者提出了第五代评价理论。该理论从实证主义转向社会建构主义,核心理念是通过行动研究使评价、决策、行动融为一体^⑨。我国学者也提出当前教育评价迎来了“以育人为本、以服务为导向、智能化为特征”的第五代评价^⑩。《深化新时代教育评价改革总体方案》明确指出,“坚持科学有效,改进结果评价,强化过程评价,探索增值评价,健全综合评价,充分利用信息技术,提高教育评价的科学性、专业性、客观性”^⑪。全球教育评价改革对传统评价观念进行了更新,为我们提供了重要的理论启示:一是评价目标从判断走向诊断,旨在促进评价对象的全面发展,并为教育决策服务;二是评价主体从评价者转向评价利益相关者共同参与,构成“评价共同体”;三是评价内容更加丰富全面,重视学生的能力提升和品德塑造,注重培养学生的国家认同感、社会责任感、综合素养和国际视野;四是评价手段趋向智能,利用人工智能、大数据等技术,开展动态化、全景式的数据采集和测量,注重终结性评价与过程性评价相结合、人工评价与机器评价相结合。

(二)教材观

① Tom Hutchinson, Alan Waters, *English for Specific Purposes: A Learning-centred Approach* (Cambridge: Cambridge University Press, 1987), 96-105.

② Michael P. Breen, Christopher N. Candlin, “Which materials?: A consumer’s and designer’s guide,” in *ELT Textbooks and Materials: Problems in Evaluation and Development*, ed. Leslie E. Sheldon (Modern English Publications in Association with The British Council, 1987), 13-28.

③ Jo McDonough, Christopher Shaw, Hitomi Masuhara, *Materials and Methods in ELT: A Teacher’s Guide*, 3rd Edition (MA: John Wiley Sons, Inc, 2003), 50-62.

④ Brian Tomlinson, “Materials evaluation,” in *Developing Materials for Language Teaching*, ed. Brian Tomlinson (London: Continuum, 2003), 15-36.

⑤ Jayakaran Mukundan, Touran Ahour, “A Review of Textbook Evaluation Checklists Across Four Decades (1970-2007),” in *Research for Materials Development in Language Learning: Evidence for Best Practice*, eds. Brian Tomlinson, Hitomi Masuhara (London: Bloomsbury Academic, 2010), 336-352.

⑥ Mohamad Syafiq Ya Shak, Intan Safinas Mohd Ariff Albakri, Siti Shuhaida Shukor, Mohd Haniff Mohd Tahir, “Criteria in English Language Textbook Evaluation Checklists: A Systematic Review,” *IARTEM.e-journal* 14, no. 1 (December 2022): 1-19.

⑦ 文秋芳《“产出导向法”教学材料使用与评价理论框架》,《中国外语教育》2017年第2期,第17—23页;赵金铭《论对外汉语教材评估》,《语言教学与研究》1998年第3期,第4—19页。

⑧ 埃贡·G·古贝、伊冯娜·S·林肯《第四代评估》,秦霖等译,中国人民大学出版社2008年版,第1—23页。

⑨ Naglaa Fathy El Dessouky, “Public Policy Evaluation Theory: From First to Fifth Generation,” *EPRA International Journal of Economics and Business Review* 4, no. 4 (April 2016): 15-25.

⑩ 刘云生《抢占教育智能化评估的制高点》,《教育发展研究》2019年第3期,第3页。

⑪ 《中共中央 国务院印发〈深化新时代教育评价改革总体方案〉》,《中华人民共和国国务院公报》2020年第30号,第11页。

教材观是对教材本质、属性和作用的认识。教材观又可细分为教材本体观和教材功能观。教材本体观侧重对教材本质和属性的看法,而教材功能观偏向对教材作用和价值理解,教材本体观与功能观具有对应关系,即教材具有什么属性,就会衍生出哪些功能。教材本体理论认为,教材具有多维属性。其中教育属性、教学属性、内容属性、物理属性受到广泛认可^①。教材的教育属性要求人们关注“培养什么样的人、如何培养人、为谁培养人”^②。教材集中体现了国家对教育教学的总体要求,影响着学生世界观、人生观、价值观的塑造,发挥着立德树人的教育功能。教学属性是教材的根本属性,规范并制约着其他属性的表达。教学属性是指教材编写者以课程标准为基本依据,通过学科知识的教学论转换,使其文本具有教师能教会教、学生易学乐学的特点^③。教学属性在教材本体中的核心地位使得教学功能亦成为教学系统中的主要功能,具体体现在它能实现教学目标、落实课程规划、保障教学质量、促进学生全面发展、支持教师教学和职业成长等方面。内容属性呼唤人们思考什么知识最有价值。教材以结构化、系统化的方式集中呈现了提纯萃取后的学科知识,发挥着传授知识和技能的功能,回答一个学科“教什么、学什么、教多少、学多少、怎么教、怎么学”的重要问题。物理属性反映在教材的插图选配、版式设计、出版质量、资源链接等外在形式。教材的物化形态使之具有保存、积淀、传承知识的功能,以及实用和商业的价值。

三 动态评价模型构建

(一) 评价理念

评价理念是建构评价模型的起点,反映了建构者对评价本质的认识,以及对评价主客体关系、价值主客体关系的理解,决定了评价模型构建过程中的目标设定、内容设计、方法选择、技术路径等,为评价指标筛选、修改等具体行为提供宏观指导。

第五代评价观和经典教材观共同构筑了评价模型的理论基石。构建评价模型应在顺应全球教育评价改革潮流、体现教材本质的基础上,进一步凸显国际中文教育的学科特色,解决当前部分国际中文教材适用性不强、内容陈旧、精品缺乏、配套不足等问题^④。教材本体理论认为,教材属性在不同类型的教材中有强弱之分^⑤。国际中文教材是以国际学生为教学对象,以中文和中华文化为核心教学内容,以培养国际学生中文交际能力和跨文化交际能力为主要教学目标的教學用书,具有第二语言教材的性质。国际中文教材除了具有教育性、教学性、内容性和物理性以外,更加凸显实践性、民族性和国际性。国际中文教材强调实践性,其编写主旨是以中文知识和技能训练为基础,以提高国际学生的语言综合运用能力为目标,具有为国际学生营造语境、提供语言交流机会,促进协商互动的实践性功能。国际中文教材反映了中华民族的语言和文化特色,以及民族历史和民族心理的样态,具有广泛而深刻的民族色彩^⑥。同时因其使用者是外国人,而蕴涵了国际性,是一种跨国籍、跨语言、跨文化的外向型教材。国际中文教材在民族性与国际性的共同影响下,承载着向世界传播中华语言文字和文化的使命,这就要求教材编写既要突出汉语汉字和中华文化的特点,又要充分考虑教材的使用环境和国际学生的需求,以跨语言、跨文化的视角,用国际学生可接受的表达方式展示教学内容。语言与文化的国际传播功能在我国主导研发的国际中文教材中显得尤为重要。综上所述,本研究将从6个方面具体阐述评价模型的理念。

1. 中外融通

教材应坚守中文和中华文化的主体性,体现中国和中华民族的基本价值观,突出中文特点,提炼展示中华文化精髓。同时,应坚持人类共同价值观,注重世界文化的多样性,尊重各国法律法规、文化传统和教育理念,促进国际学生对中国的认知、理解乃至认同。

① 曾天山《教材论》,人民教育出版社2019年版,第12—15页。

② 《习近平在全国高校思想政治工作会议上强调 把思想政治工作贯穿教育教学全过程 开创我国高等教育事业发展新局面》,《人民日报》2016年12月9日,第1版。

③ 王攀峰、宋雅琴《论教科书的内涵与属性》,《当代教育科学》2018年第1期,第7—12页。

④ 教育部中外语言交流合作中心《国际中文教育教学资源建设行动计划(2021—2025年)》,2021年12月, <http://www.chinese.cn/uploads/file/20220125-1643091013876953.pdf>。

⑤ 曾天山《教材论》,人民教育出版社2019年版,第12—15页。

⑥ 李泉《对外汉语教材通论》,商务印书馆2012年版,第8—9页。

2. 素养提升

教材应立足学生的全面发展。内容涵盖中文知识与技能、中国文化与当代国情、情感态度、语言学习策略和人类共同价值观塑造,促进国际学生中文综合运用能力、跨文化交际能力、学习能力、国际理解能力的全面提升。

3. 教学适用

教材应注重适用性,贴近教学实际情况,符合学生学习特点和文化背景,助力教师组织开展教学,实现教师专业发展。同时,应遵循中文作为二语教学规律,体现二语教学方法,提供清晰的二语教学思路,推动教学模式创新,提高教学效率。

4. 应用实践

教材应强调语言文化知识和技能的综合运用,为汉字、词汇、语法提供恰当的语境,选择符合交际需要的主题和内容,强化文化的体验、比较和思辨,以丰富多样的实践活动贯穿教材的学习内容,强调活动的真实性、互动性、有效性、综合性和可操作性。

5. 科学规范

教材应重视知识体系的科学性和规范化。语言知识体系以《国际中文教育中文水平等级标准》(GF0025—2021)为依据,对应“三等九级”应掌握的音节、汉字、词汇、语法量化指标,确保教材中的语言规范,知识分布合理。文化知识体系以《国际中文教育用中国文化和国情教学参考框架》为依据,对应中国社会生活文化、中国传统文化、当代中国国情三大板块的文化项目,兼顾古今,立体呈现。教学内容应循序渐进,衔接紧密,难度合理上升。

6. 技术赋能

教材应在需求调研基础上配备相应的数字化教学资源。教材的评价方式应借助智能技术手段,依据评价目标、对象、内容,创新评价过程与方法,构建专业、高效、综合的评价模型,实现客观、即时、精准评价。

(二) 评价对象与目标

国际中文教材类型多样,根据研发主体可分为我国主导研发的中文教材和非我国主导研发的中文教材,根据使用环境可分为目的语环境中文教材(即来华留学生中文教材)和非目的语环境中文教材(即海外本土中文教材),根据课型又可分为综合课、技能课等不同课型中文教材。本研究将我国主导研发的国际中文综合教材作为重点评价对象,理由有三点:其一,综合教材是国际中文教材建设的重点,它是目的语环境中中文教学的主干教材,在非目的语环境中一般仅使用综合教材;其二,国际中文综合教材内容全面系统,它包括中文和中华文化的基本教学内容,涵盖听、说、读、写、译五项语言技能,涉及语音、汉字、词汇、语法四项语言要素;其三,综合教材具有较强的普适性,以此为基准构建评价模型,对推动国际中文教材建设,具有更为普遍的指导意义。

构建评价模型的总体目标是“以评促建、以评提质”。该模型基于“国际中文教材动态评价系统”,支持开展自编教材或商业教材的网络实时诊断性评价及动态质量监测,强化服务导向,可以帮助海内外中文教学机构、中文教师和学生甄选教材,支持国际中文教师优化教材使用策略,实现个性化的教学实践,促进教师专业能力建设,引导国际中文教材研发机构和编写者编写、改进教材,提高教材编写质量,服务相关管理机构开展基于客观数据的国际中文教材认证与推荐。

(三) 评价内容

评价模型由规范评价、定量评价、定性评价和附加评价四部分构成,依据教材的教育属性、教学属性、内容属性、物理属性划分为价值取向、教学适用、内容编排、外观配置4个一级指标,下设20个二级指标,60个三级指标。

1. 第一部分:规范评价

规范评价是国际中文教材必须达到的要求,包括3个二级指标,6个三级指标。规范评价只有“合格”和“不合格”两个选项,若有一项指标为“不合格”,评价就此终止,教材将被判定为“未达到标准”。若上述指标均为“合格”,评价继续进行。

表 1 国际中文教材规范评价表

一级指标	二级指标	三级指标
(一)价值取向	1.国家意识	(1)体现中国和中华民族的基本价值观
		(2)符合中国及各国法律法规
		(3)维护中国及各国主权和领土完整
	2.世界认知	(4)尊重各国文化传统和风俗习惯
		(5)没有种族、地域等方面的歧视和偏见
	3.规范要求	(6)没有版权争议,不侵犯其他个人或组织的著作权

2.第二部分:定量评价

定量评价是对国际中文教材中的汉字、词汇、语法等的量化评价,包括 2 个二级指标,10 个三级指标。定量评价由“国际中文教材动态评价系统”自动判定,若指标数值在参考区间内,该指标为“合格”,若指标数值不在参考区间内,该指标为“不合格”。系统完成定量分析后,评价继续进行。

表 2 国际中文教材定量评价表

一级指标	二级指标	三级指标
(二)内容编排	4.语言知识	(7)汉字等级分布
		(8)汉字编入比例
		(9)词汇等级分布
		(10)词汇编入比例
		(11)生词的复现率
		(12)平均每课后语法数
		(13)平均每课生词数
		(14)平均每课生词密度
	5.课文	(15)课文句子总数
		(16)课文平均句长

3.第三部分:定性评价

定性评价是对国际中文教材编写质量的通用评价,包括 12 个二级指标,37 个三级指标。每个指标按 1—5 评分,1=非常差,2=差,3=一般,4=好,5=非常好。若指标不适合评价目标教材,可选择 NO。若目标教材为通用中文教材,评价就此终止。若目标教材为海外本土中文教材,评价继续进行。

表 3 国际中文教材定性评价表

一级指标 (权重)	二级指标 (权重)	三级指标
(三)教学适用 (20%)	6.课程适用 (10%)	(17)教学目标清晰明确,参考现行的课程标准或教学大纲确定教材内容
		(18)提供课时建议,课时安排合理
	7.学生适用 (6%)	(19)符合学生的中文水平,难易适中
		(20)符合学生的年龄和认知水平
		(21)符合学生的学习目标 and 需求
		(22)学习内容能引发学生的兴趣
	8.教师适用 (4%)	(23)有利于引导教师备课,便于教师使用
		(24)赋予教师灵活使用的空间
		(25)有利于教师对教学效果进行评估

(四)内容编排 70%	9.编写理念 (10%)	(26)合理吸收并体现相关语言教学理论的研究成果,并符合第二语言教与学的规律
		(27)教材整体或局部设计有创新、有特色
	10.框架结构 (4%)	(28)每单元/课中各个模块设计合理,相互联系
		(29)教学内容(包括语言知识、语言技能、文化等)布局合理,联系紧密,难度递增适度
	11.主题内容 (16%)	(30)主题/话题丰富多样,符合学生真实交际的需要
		(31)主题/话题编排由易到难,循序渐进,符合教学规律
	12.语言内容 (16%)	(32)语言真实、自然,具有交际性
		(33)语言规范、准确
		(34)为生词和语法提供了典型例句,必要时提供了上下文,使语境更明确
		(35)生词和语法的注释和翻译简明准确,易于理解
		(36)听、说、读、写、译各项技能训练比重合理
	13.文化内容 (10%)	(37)包含中国社会与人民生活的文化内容
		(38)包含中国优秀传统文化的文化内容
		(39)包含当代中国国情的文化内容
		(40)包含与语言使用和交际相关的文化因素
		(41)文化内容的编排合理适度,呈现方式自然恰当
	14.活动设计 (10%)	(42)活动/练习的类型和形式丰富多样
		(43)活动/练习有利于学生有效地掌握所学内容
		(44)活动/练习能促进有意义的互动
		(45)提供综合运用中文技能的活动/练习
		(46)活动/练习目标明确
		(47)活动/练习指令清晰、简洁,可操作性强
	15.策略培养 (4%)	(48)提供主要的语言学习策略(如推测、联想、自评、查找资料等),帮助学生更好地掌握学习内容
		(49)提供积极的交际策略(如解释、更正、重复、澄清、请求等),帮助学生更好地完成交际任务
(五)外观配置 (10%)	16.形式设计 (5%)	(50)版式设计(如字体字号、中外文及拼音混排、布局等)清晰美观,具有功能性
		(51)插图与内容配合紧密,具有功能性,体现时代感和中国特色
	17.配套资源 (5%)	(52)根据教学需要,配有必要的辅助材料,如练习册、测试题、教师用书、字词卡片、挂图等
		(53)根据教学需要,配有数字化教学资源,如视频、配套网站、App等

4.第四部分:附加评价

海外本土中文教材指专为特定非目的语国家或地区的中文学习者编写的国际中文教材。编写方式主要分为三种:中方主导编写、外方主导编写和中外合作编写。附加评价是对海外本土中文教材的追加评价,包括3个二级指标,7个三级指标。每个指标按1—5评分,1=非常差,2=差,3=一般,4=好,5=非常好。若指标不适合评价目标教材,可选择NO。

表 4 国际中文教材附加评价

一级指标 (权重)	二级指标 (权重)	三级指标
(六)教学适用 (20%)	18.本土化教学环境 (20%)	(54)符合当地主流外语教学理念
		(55)符合当地课程设置
(七)内容编排 (70%)	19.本土化内容 (70%)	(56)包含贴近当地生活的主题内容
		(57)体现当地学生在中文语言要素方面的学习难点
		(58)融入当地文化的内容,恰当处理文化比较中的异同
		(59)用当地语言注释,且简洁、准确、易懂
(八)外观配置 (10%)	20.本土化设计 (10%)	(60)版式和插图设计体现当地文化特色

(四)评价方式

1.机器评价与人工评价相结合

以往的教材评价主要由定性描述的指标构成,以教师或专家为评价主体,即使采用评分量表的形式,主观性仍较强^①。本研究探索了机器评价的可行性。在规范评价部分,系统利用敏感词提取技术,为评价者提供智能提示结果,评价者将其逐一审核后,可自动引用某些结果作为评价证据。在定量评价部分,系统可自动识别和分析教材中的汉字、词汇、语法、课文等 10 个定量指标所涉及的内容,对比参数后,分别作出合格/不合格的可视化评价。机器评价结果可为人工评价提供客观的数据参考。评价者在系统中根据定性评价和附加评价的指标要求,结合教材内容进行逐一打分,及时记录在教材中发现的相关证据,完成人工评价任务。整个评价流程基于网络平台,适于自评或他评,能够较好地将定量评价与定性评价相结合,利于实施多元主体评价,可显著提高评价的准确性和客观性。

2.诊断评价与循证评价相结合

传统的教材评价侧重结果,仅呈现评分或者等级,存在理据性不足、指导性欠缺等问题。本研究建构的评价模型可用于自编讲义或公开出版教材,可用于教材编写过程中或编写完成后,也可用于教材的使用前、使用中或使用后。评价者可按提示步骤进行教材实时评价,评价结束时可获得一份诊断报告,报告内容包括教材评价总分值、每个维度的分值、具体情况以及修改建议等。即使是定性评价指标,评价者也要提供充分的证据,如:“第×页,练习×,练习设计的目标不够明确,建议改为……。”在评价过程中充分循证,给予及时诊断,可为编写者或使用用户提供建设性意见,有助于改进教材编写,调整教材使用策略。

(五)构建方法

评价模型的构建与完善,采用了多种研究方法,主要分为以下步骤。

1.指标描述语库的建立与应用

搜集中英二语教材评价量表共 57 个,建立指标描述语库(含 2221 个指标),通过词频统计、多词序列提取、搭配分析 3 种语料库分析手段,构建国际中文教材评价指标基础框架。

2.一线教师和专家的团体访谈

采用焦点团体访谈法,分别对 8 位一线资深中文教师和 7 位业内专家进行了两次访谈,对评价基础框架中的指标进行了修订和完善。两次访谈确立了国际中文教材评价指标框架,在框架中增加了价值取向维度,区分了通用教材评价指标和本土教材评价指标,区分了定性评价指标和定量评价指标。

3.第一次德尔菲法专家咨询

对 13 位业内专家进行了匿名咨询,旨在进一步修改并精简指标框架。经过三轮咨询,专家意见趋于统一,国际中文教材评价指标体系初步构成。

^①梁宇《教师为评价者的国际汉语教材评价标准实证研究》,《国际汉语教育(中英文)》2017 年第 3 期,第 48—58 页。

4. 问卷调查与因子分析

面向国内外中文教师进行了较大规模的问卷调查,共回收有效问卷 1002 份,其中 501 份用于探索性因子分析,501 份用于验证性因子分析。最终收敛成 8 个显著因子,经过此步骤,评价指标体系得到了进一步优化。

5. 第二次德尔非法专家咨询与权重分配

邀请了 8 位业内专家,将上述步骤的研究成果发给专家,并详细说明咨询目的和咨询问题。通过三轮咨询,运用层次分析法得出了定性评价和附加评价中 4 个一级指标和 15 个二级指标的权重值,最终确立了国际中文教材评价指标体系。

6. 评价模型与支撑平台搭建

通过对 3000 册国际中文教材的语言要素进行统计分析,建立等级词汇库、汉字库、话题库、文化点库、年度热词库、敏感词库等知识库,实现对国际中文教材的定性和定量评价。利用自然语言处理技术、深度学习技术和可视化技术,创建国际中文教材评价算法,以此为平台核心,并通过微服务、分布式、容器化、DevOps 和持续交付等技术构建了高效可靠的应用平台。

四 动态评价模型特色

(一) 评价内容特色

评价模型的主体内容采用“四维”架构,其中价值取向考察教材的总体导向,教学适用观察教材的基本特性,内容编排评估教材的基础要素,外观配置考量教材的外在条件。“四维”架构的评价内容还具有如下特点。

1. 坚守中华文化立场,打造柔性话语表达

党的二十大报告中提出,“坚守中华文化立场,加快构建中国话语和中国叙事体系”^①。外语教材本身承载着价值和意识形态^②。我国主导研发的国际中文教材既是对外展示传播我国语言文化的重要窗口,更是坚守中华文化立场、彰显中华民族价值观的重要场域。对其评价,首先要衡量它是否符合我国主流的价值规范。这在实际的教材编写和选用过程中已成为约定俗成的评价准则,却是首次作为正式指标纳入评价模型,我们尝试回答“什么是国际中文教材正确的价值取向”,其中尤为重要的即是如何看待“我者”和“他者”。对“我者”,应坚定语言自信和文化自信,恪守中华民族基本价值观,遵守国家法律法规,维护国家主权和领土完整。三者在评价模型中凝聚为“国家意识”维度,在首轮咨询和调查中受到认可,权重居二级指标之首,足见其重要性。同时,在法律法规和主权领土两个指标描述中以“我国及各国”为定语,体现了关照“他者”的态度。另外在“世界认知”、“本土化教学环境”、“本土化内容”、“本土化设计”中,重点体现了尊重与包容“他者”文化和感受的国际主义精神,这种精神表征实际源于更深层次的文化底蕴和自信。从关怀“他者”的视角出发,要对教材的表达方式加以考量,要在教学内容上与他国的语言特点和社会文化进行比较与融通,在话语实践上考虑他国师生的理解和接受方式,多利用他国学生易听、易懂、易记,能激起共同认识和情感共鸣的主题内容和表现形式^③。其次,恪守中华文化主体性是坚守中华文化立场的核心要义。评价模型中专设“文化内容”维度,其具体指标对应《国际中文教育用中国文化和国情教学参考框架》中的中国社会生活文化、中国传统文化、当代中国国情三大板块,考察教材中中华文化内容的选择与呈现,重点关注了教材如何处理传统文化与现代文化、知识文化与交际文化、中华文化与母语文化的关系,突出了语言中的文化,体现了以培养国际学生跨文化交际能力为目标的文化教学理念和文化内容编写理念。

2. 注重两种能力培养,聚焦中文二语教学

国际中文教育必须完成好培养对象国学生具有“跨语言交际”和“跨文化交际”两种交际能力^④。两种交际能力相辅相成,交融互动,协同发展,犹如语言与文化密不可分。评价模型将跨语言交际能力显性地表现在语言知识、语言技能、主题、语篇、语言活动、策略培养等指标。这部分的指标数量占比约 50%,体现了评价模型以

① 习近平《高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告》,第 45—46 页。

② Xiao Lan Curdt-Christiansen, Csilla Weninger, “Ideology and the Politics of Language Textbooks,” in *Language, Ideology and Education: The Politics of Textbooks in Language Education*, eds. Xiao Lan Curdt-Christiansen, Csilla Weninger (London: Routledge, 2015), 1.

③ 梁宇《国际中文教材国家形象自塑的多元表达》,《云南师范大学学报(哲学社会科学版)》2022 年第 3 期,第 57—65 页。

④ 陆俭明《新时代国际中文教育理念创新和实践探索的若干思考》,《语言教学与研究》2022 年第 4 期,第 1—8 页。

中文作为第二语言教学为重心的基本立足点。在中文交际能力培养方面,评价内容突出了以下4个特点。

(1)以主题为统领。评价模型中主题设计置于语言、文化、活动内容之前,发挥其统领教学内容的先导作用,侧重对其丰富性、交际性和循序渐进性的考察。正如 Curtain & Dahlberg 所述,语言教学中的主题设计在连接内容、语言和文化目标等方面具有优势,还有利于创设情境、建构意义、关联现实生活、强调真实语用及培养高级思维^①。

(2)以语言知识为基础。语篇具有承载语言知识的功能,教材评价应突出语篇的多样性、真实性和交际性。体系中评价语言知识的指标数量最多,并采用了定量定性相结合的评价方式,就是要强调语言知识在发展中文交际能力方面的基础作用。

(3)以实践活动为主线。Tomlinson 指出,好的教材应为学生创造运用语言进行真实交际的机会,这种机会是通过大量的互动性、有意义的活动实现的^②。评价模型中的活动采用了广义概念,将教材中所有的语言实践项目纳入评价范畴,涵盖了练习、任务、游戏等形式,拓展了活动的内涵。评价模型对活动设计的评量尤为细致,不仅凸显了活动在促进有意义的互动方面实现的语言交际价值,还依据多样性、有效性、综合性、目的性、易操作性等活动设计原则给出了“好的语言活动”的标准。

(4)以策略培养为辅助。策略作为“内隐的学习规则系统”,对知识和技能教学具有隐性的推动作用,是促进知识学习、增强教学效果的底层动力。虽然策略在教学中日益受到关注,却往往在教材编写和评价中被忽略。评价模型专设“策略培养”维度,从主要语言策略和积极交际策略两方面进行构建,体现了评价模型对培养学生学习能力的关注。

在跨文化交际能力方面,评价模型具体表现在价值取向、文化知识、文化呈现、文化态度等维度,以及附加评价中,前文已有论述,此处不再赘言。

3.突出教学适用评价,深化教育服务理念

“没有完美的教材,只有合适的教材。”这是二语教材评价领域的经典名句,体现出教材“适者生存、适者为优”的生态学观点,由此“适用与不适用”逐渐替代“好与不好”,成为二语教材评价中更加贴近教学实际、目标更为具体、操作性更强的重要标准。二语教材评价有重视适用评价的传统,但所涉维度略有不同。Grant 评价量表由三个适用维度构成:教材符合学生需要、教材符合教师需要、教材符合课程大纲和考试要求^③。Cunningsworth 提出教材的适用评价主要衡量教材与特定要求的匹配度,如学生的学习目标、学习背景等^④。Byrd & Celce-Murcia 建构了教材与课程、教材与学生、教材与教师的适用评价量表^⑤。评价模型秉承二语教材适用评价的理念和主要观测角度,设计了教学环境适用、课程适用、学生适用、教师适用“四个适用”的评价框架。教材是一定教育文化环境中的产物,教材直接体现课程规定的目标及内容,教材要落实“以学生为中心”的教学原则,注重学生中文水平、认知水平等个体因素变量,教材辅助教师教学、促进教师专业发展。可见,教材与环境、课程、师生具有紧密的互动关系,而适用评价则是对这种互动关系最为恰切的评价。“教学环境适用”专为评价本土中文教材而设计,因本土教材使用环境的特殊性,考察教材在他国教育文化中的适用性显得异常重要。“四个适用”的评价为当前国际中文教材不适用的现实问题提供了改进思路。适用评价表明教材评价不仅关注知识和内容,更看重教材与教学要素之间的互动,紧扣教材教学论的基本原理,强化落实教材服务教学的核心宗旨。

(二)评价方式特色

从评价方式角度,评价模型在开展综合性、过程性和发展性评价方面进行了创新性探索。

① Helena Anderson Curtain, Carol Ann Dahlberg, *Languages and Learners: Making the Match: World Language Instruction in K-8 Classrooms and Beyond*, 5th Edition (London: Pearson Education, 2014), 16.

② Brian Tomlinson, “Introduction: Principles and Procedures of Materials Development,” in *Materials Development in Language Teaching*, 2nd edition, ed. Brian Tomlinson (Cambridge: Cambridge University Press, 2011), 1-34.

③ Neville Grant, *Making the Most of Your Textbook* (Essex: Longman, 1987), 118-121.

④ Alan Cunningsworth, *Choosing Your Coursebook* (Oxford: Macmillan Education, 1995), 6.

⑤ Patricia Byrd, “Textbooks: Evaluation for Selection and Analysis for Implementation,” in *Teaching English as a Second or Foreign Language*, 3rd Edition, ed. Marianne Celce-Murcia (MA: Heinle & Heinle, 2001), 415-428.

1. 引入链式思维, 强化综合性评价

链式思维源于管理学, 基于系统论和全局观, 主张每个事物都像链条上的一个环节, 环环相扣, 前后关联, 形成整体。链式思维注重明暗线结合、分工明确、结构严谨, 这与强调评价对象整体性、系统性的综合评价理念非常契合。国际中文教材是一个复杂精密的系统, 既有宏观理念的考量, 也有微观知识点的选择和排布, 既有外部教学因素的影响, 更看重内容的表达和逻辑, 既强调语言技能的实践, 又要体现多元能力的培养。构建评价模型, 对其进行系统性、综合性评价需要借用链式思维方式。在评价模型中, 规范评价、定量评价、定性评价、附加评价可视为评价链上的四大节点, 各级指标为分节点, 评价实践按步骤、分节点有序展开, 先进行规范评价, 考察教材是否达到最基本的要求, 再通过定量评价的客观数据辅助其后的主观评价, 先评价通用内容, 后追评本土内容, 兼顾通用教材评价和本土教材评价, 拓展了评价对象的范围。

2. 强调动态开放, 践行过程性评价

在教材评价领域, 过程评价是指对教材编写、使用、改进的全过程进行评价和监测。本研究构建的评价模型, 借助多种技术手段, 能够对教材进行多维度的自动分析, 并将机器评价与人工评价相结合, 实现基于客观数据的主观评价, 提升教材评价的便捷性和客观性。评价模型注重多元数据的采集与分析, 包括机器自动生成的字、词、句、语法的统计结果, 以及人工在各指标上的评分和证据描述等, 这种多元数据采集方式可有效提升评价结果的准确性和精细度。该模型鼓励多元主体参与评价, 评价者能够实时接收评价反馈, 多元视角的评价反馈汇聚成大数据, 有助于提升教材评价的精准性, 还能够实现教材的智能推荐, 为教材选用提供便利。

3. 重视诊断反馈, 实现发展性评价

教材的发展性评价是指发挥教材评价诊断、调节、反馈、激励的作用, 达到促进教材发展的目标。该模型能够生成可视化的评价诊断报告, 其中包含丰富、具体且富有建设性的评价意见, 包含量化统计结果和参考区间值, 与相关标准比对后未编入教材和超纲的字、词、语法, 各指标的评分及依据、修改建议等。评价的诊断和反馈在教材编写和教材使用过程中发挥着调节和促进作用。在教材编写过程中, 评价的诊断和即时反馈有益于教材改进, 提升教材编写质量; 在教材使用过程中, 诊断和反馈能够为教师提供教材调整和改编策略, 以提高教材的使用效率。

五 结语

教材评价是提升教材质量的有力抓手, 但国际中文教材涉及内外变量众多, 对其评价也是教材研究的难点。本研究综合运用了语料库、专家咨询、问卷调查等研究方法, 借助自然语言处理、深度学习等技术手段, 尝试从评价理念、目标、对象、内容、方式等方面系统构建评价模型, 力求在评价内容方面突出时代特色和学科特色, 在评价方式方面践行综合评价、过程评价和发展评价。未来研究亟需开展国际中文教材评价实践, 以检验评价模型的有效性, 并通过教材评价数据采集, 测算各个评价指标的及格水平参考值, 进一步优化评价模型并增强评价模型的科学性和适用性。

[责任编辑: 唐 普]