

基于增值评价的“三维”大学外语综合型语言教学评价模型构建研究

杨 雪

武汉纺织大学外国语学院 湖北武汉

【摘要】本研究尝试构建基于增值评价的“三维”大学外语综合型语言教学评价模型。模型通过整合增值评价 DEA 模型、SGP 模型与过程评价和结果评价，形成综合型的语言教学评价体系。同时基于模型实施了为期 1 年的实证研究。结果显示，实验组各项语言技能的增值幅度均显著大于对照组，验证了模型对教学质量提升的积极作用。该模型突破了传统评价的单一性局限，实现了从“静态结果导向”向“动态发展导向”的转变，为大学外语教学提供了科学、可操作的评价工具与“评价--教学”的互动机制。

【关键词】语言教学评价体系；增值评价；DEA 模型；SGP 模型

【基金项目】第十二批“中国外语教育基金”项目（项目编号 ZGWYJYJJ12A166）：大学外语课堂“三维”综合型语言评价体系模型构建

【收稿日期】2025 年 10 月 6 日

【出刊日期】2025 年 11 月 8 日

【DOI】10.12208/j.ije.20250405

Research on the construction of a "Three-Dimensional" comprehensive language teaching assessing model for college foreign languages based on value-added assessment

Xue Yang

Wuhan Textile University, School of Foreign Languages, Wuhan, Hubei

【Abstract】 This study attempts to construct a "Three-Dimensional" comprehensive language teaching assessing model for college foreign languages based on value-added assessment. The model integrates the value-added DEA model, SGP model, process evaluation and outcome evaluation to form a comprehensive language teaching assessing system. At the same time, an empirical study was carried out for one year based on the model. The results show that the value-added amplitude of each language skill in the experimental group is significantly greater than that in the control group, verifying the positive role of the model in improving teaching quality. This model breaks through the limitation of the singleness of traditional evaluation and realizes the transformation from "static result-oriented" to "dynamic development-oriented", providing a scientific and operable assessing tool and "assessment-teaching" interaction mechanism for college foreign language teaching.

【Keywords】 Language teaching assessing system; Value-added assessment; DEA model; SGP model

1 研究背景及目的

随着全球化进程的加速，外语能力已成为衡量个人竞争力的重要指标之一。大学外语教学作为培养具备国际视野人才的关键环节，对其教学效果的评价显得尤为重要。2020 年国务院印发的《深化新时代教育评价改革总体方案》明确提出“改进结果评价，强化过程评价，探索增值评价，健全综合评价”的指导方针，强调评价体系需从“静态结果导向”转向“动态发展导向”。同年发布的《大学英语指南（2020 版）》进一步指出，大学外语评价应“关注学生学习过程中的进步与变化，构建多元评价体系以促进语言能力发展”。政

策层面的顶层设计为外语教学评价改革提供了明确的方向：即通过增值评价挖掘学生个体潜力，通过过程评价诊断学习问题，通过结果评价检验发展成效^[1,2]。

当前大学外语评价方式基本以过程评价、结果评价为主。但普遍存在如个体差异被忽视、反馈改进脱节等痛点。过程评价注重学生在学习过程中的参与度和活动表现，强调在真实情境下的语言实践。结果评价通过对最终学习成果的评估，直观反馈教学效果。增值评价作为一种关注学生“净增值”的纵向评价方法，强调学生学习的进步幅度，关注在一定时间内的成长变化，突出了评价的过程性和连续性。三种评价模式各有所长，

为了实现更为科学、合理的教学评价,可以尝试将三种评价模式结合,形成一个有机的整体,也许能成为破解传统评价局限的有效路径。

本研究尝试通过构建“增值评价为框架、过程评价为支撑、结果评价为验证”的三维理论模型,开发适用于大学外语课堂的可操作化评价工具,包基于增值评价 DEA 模型的过程评价指标体系和基于增值评价 SGP 模型的结果评价指标体系,通过实证研究验证模型对教学质量提升的作用,希望能够形成可推广的“评价—教学”互动机制。

2 评价理论及国内外研究现状

2.1 评价理论

增值评价 (Value-Added Assessment) 是一种纵向评价方法,强调以学生个体为基准,通过纵向数据对比,量化其在知识、能力、素养等方面的进步幅度,即比较学生在特定时间段内的实际表现与预期表现,测量其“净增值”(排除学生原有水平、家庭背景等因素后的真实进步)。其核心价值在于关注个体差异,强调“进步即成功”。其理论基础包括多元智能理论(关注个体差异)、学习投入理论(强调过程性成长),并通过具体模型实现量化分析,本研究主要选用了 DEA 模型(数据包络分析)和 SGP 模型(成长百分位模型)。DEA 模型源于运筹学,通过构建“投入—产出”指标体系,评估决策单元(如学生个体)的相对效率。在本研究中,投入指标包括学习态度、学习策略及学习时间等,产出指标包括语言知识、语言技能、及情感能力的增值得分;SGP 模型通过将学生成绩与相似能力水平的同伴群体

对比,计算其在群体中的成长百分位,直观反映个体进步的相对性。过程评价 (Process Assessment) 聚焦学习过程中的行为、策略与态度,通过课堂观察、学习日志、同伴互评等方式收集数据,强调形成性反馈对学习的促进作用。结果评价 (Product Assessment) 关注学习成果的终结性测量,以标准化测试、项目成果展示等为主要形式,强调目标达成度与标准参照。可见三种评价方式在评价的过程、评价数据和结果反馈方面是有一定交集的,可以各取所长,组建一套综合型的评价体系。

2.2 国内外研究现状

在增值评价方面,英国自 2002 年起将“增值分数”(Value Added Score)纳入学校效能评估体系,用于衡量学校对学生进步的贡献^[3];美国田纳西州的 TVAAS 系统通过 SGP 模型评估教师效能,成为增值评价在基础教育领域的经典案例^[4]。我国学者近年来对于增值评价在教育领域的应用也开展了一定的研究。并指出我国增值评价面临“数据质量、模型适配、文化差异”三大挑战^[5,6],提出“评价目标与教育目标协同”的本土化原则^[7];有研究者将 OBE 理念与增值评价结合^[8,9],尝试基于增值评价的混合式的教学模式^[10],并构建系统性的^[11]生态性的^[12]教学质量评价体系,为外语课程改革提供了新思路。

通过在中国知网 (CNKI) 以“增值评价”“英语教学评价”等为核心关键词进行共现情况检索,系统梳理该领域的研究热点与前沿动态。如图 1 所示,借助 CiteSpace 的关键词突现分析可知:2023 年后“增值评价”的相关研究呈上升趋势。



图 1 2015-2025 年增值评价相关研究关键词突现强度 TOP15

总体而言,当前对于增值评价在语言教学评价方面的研究相对较少,缺乏对增值、过程、结果评价的系统性整合研究,总体呈现外语学科特性不足。

3 “三维”大学外语综合型语言教学评价模型构建

3.1 增值评价模型选择

3.1.1 DEA 模型

增值评价 DEA 模型通过“投入—产出”分析框架,量化学习过程的效率增值,识别“高投入低产出”或“低投入高产出”的学习行为模式。过程评价核心目标为跟踪学习过程中的行为、策略与资源利用,可为增值分析提供微观证据,形成“行为投入→效率产出→增值诊断”的闭环。其指标体系设计可依据目的导向、因果关联、动态可测、可操作性原则。即围绕大学英语教学目标(知识目标、能力目标、情感目标),筛选与增值目标直接相关的输入与输出指标,剔除冗余变量。投入指标反映学习过程中的资源消耗(如时间、精力),产出指标体现过程性成果(如知识掌握、技能提升);指标需可量化、可追踪,支持 DEA 模型的效率计算(如周均学习时长、课堂互动频次);聚焦学生可自主调控的行为(如学习策略选择、课堂参与度),而非不可控因素(如家庭背景)。指标需量化且便于采集,如投入指标包括自主学习时长、课堂参与度等,产出指标涵盖语言知识、语言能力及应用能力增值。

在具体 DEA 模型选择方面可使用 BCC 模型(规模报酬可变),适用于学习过程的动态性(如初期投入效率低,随策略优化效率提升),可分解纯技术效益(TE)即学习策略与规模效率(SE)即投入规模。通过获取三级指标原始分数,依据三级指标权重计算出二级指标中各项投入和产出数值,利用 SPSSAU 软件 DEA 模型,计算出相对学习效率。效率值=1:则该学生的学习过程为 DEA 有效,投入产出比最优(如“高投入高产出”或“低投入适度产出”);效率值<1:则存在投入冗余或产出不足,如“学习时间长但策略低效导致产出滞后”(需优化策略),或“课堂参与积极但资源利用不足导致效率低下”(需引导优质资源使用)。另外,鉴于 DEA—BBC 模型展示的学习效率为相对学习效率,为弥补这一不足,可以通过过程性评价的具体测量方法,以雷达图可视化展现学生在各项投入和产出与教师期待达成比率。

3.1.2 SGP 模型

增值评价 SGP 模型通过确定基线成绩、划分同质群体、计算成长百分位三个步骤,将学生成绩转化为在同质群体中的相对成长位置,直观反映个体进步幅度。

例如,若学生 SGP 值为 80,表示其成长超过了同质群体中 80%的学生。其模型设计原则遵循个体导向(以学生个体成长为核心,关注进步幅度)、多维综合(涵盖学习成绩、应用技能、综合素质等多个维度)、数据可及(指标数据易于采集和量化)。指标设计原则遵循纵向对比导向(以学生入学基线为参照,追踪其在特定周期内的成绩变化轨迹)、群体参照校准(基于同质性群体计算百分位)、多元成果覆盖原则(兼顾语言知识、技能、素养等多维目标)。因此,其指标体系应涵盖基线定位指标、成果验证指标和增值诊断指标,涉及对于学习成绩增值、应用技能提升和综合素质发展三个主要方面的测评。

3.2 “三维”语言教学综合评价体系模型框架设计

综合评价体系模型框架的构建基于评价的全面性、科学性和实用性原则。全面性即整合增值评价(关注进步)、过程评价(关注学习行为)、结果评价(关注最终成果),形成能够综合评价学习效率增值、能力成长增值、情感态度增值的“三维”评价维度。科学性即基于 DEA 与 SGP 模型量化增值,结合《大学英语指南》《中国英语能力等级量表》以及大学英语课程具体教学目标等标准设计指标,确保评价结果客观可测。实用性即通过 PDCA 循环(计划—实施—反馈—改进)实现模型动态优化,为教学提供即时反馈。

增值评价作为核心框架,串联过程评价与结果评价:过程评价为基础,提供学习行为数据,支撑增值评价的效率分析;结果评价为验证,通过最终成果检验增值目标的达成度;增值评价为纽带,量化过程与结果的关联,形成“投入—过程—产出—增值”闭环。课堂观察表与 DEA 模型投入指标联动,确保过程数据直接服务于增值计算;期末测试题目设计与 SGP 模型基线数据对标,保证成长百分位计算的科学性。

在实施流程方面,通过教学平台数据及教师课堂观察表收集学习过程、测试成绩等数据(如表 1),并运用 DEA 与 SGP 模型计算增值分数,结合过程记录与结果测试形成综合评价报告(如表 1)。教师根据评价结果调整教学策略,学生制定学习改进计划,形成“教学—评价—改进”循环。

4 实证研究

为验证模型对于大学英语教学质量的提升作用,本研究选取了武汉纺织大学 24 级大学英语 8 个班级共 275 名同学作为实践研究对象,其中有效数据 238 人。研究时长为 1 年 2 个学期。学习成绩数据包括学生分级考试成绩、2 个学期的期末考试成绩,用于评估学生

的阶段学习成果和增值情况。

首先对 8 个班的分级考试成绩进行单因素方差分析 (ANOVA)，F 值=0.48，p=0.86>0.05，8 个班成绩方差齐性，p 值均>0.05 (12401=0.58，12402=0.42，…，12408=0.63)，均符合正态分布。无极端异常值 (Z 绝对值<3)，数据质量良好，无显著差异，可以用于教学实践对比。

选取 6 个班级为实验组，2 个班级为对照组。通过对各时间点 (入校 T1、第一次期末考试 T2、第二次期末考试 T3) 分组差异检验 (独立样本 t 检验) 结果显示两组 T2 成绩无显著差异，实验组 T3 成绩接近显著高于对照组，如表 2。

同时，数据分析结果显示，实验组 Time 主效应显著 (F=125.32，p<0.001)，T2-T1 提升 24.43 分 (p<0.001)，T3-T2 回落 6.22 分 (p<0.001)；对照组 Time 主效应显著 (F=68.75，p<0.001)，T2-T1 提升 25.22 分 (p<0.001)，T3-T2 回落 6.93 分 (p<0.001)。两组均呈现“显著提升→小幅回落”的趋势 (Time 主效应 p<0.001)，对照组回落幅度 (6.93 分) 略大于实验组 (6.22 分)，但差异未达显著 (p=0.38)。

然而，从语言单项技能 (听力、语法、阅读、词汇、写作) 增值情况来看，实验组增值情况明显大于对照组 (如表 3)。

表 1 “三维”综合评价体系模型数据采集

评价维度	采集方法	技术工具	频率
SGP 模型增值评价	基线测试+期末测试	分级考试系统 (对接《量表》)	期初/期末各 1 次
DEA 模型过程评价	课堂观察+日志分析	教学平台数据分析	每单元汇总
结果评价	标准化测试	机考系统	期末集中测

表 2 T1-T3 组间差异对比

检验变量	实验组均值±SD	对照组均值±SD	t 值	df	p 值	均值差
T1	41.82±10.05	38.23±9.82	2.28	246	0.023	3.59
T2	66.25±12.58	63.45±13.21	1.36	246	0.175	2.8
T3	60.03±12.01	56.52±12.68	1.82	246	0.069	3.51
T3-T1	18.21±10.87	18.29±11.23	-0.05	246	0.96	-0.08

表 3 T1-T3 组间单项差异对比

技能	T1	T2	T3	组间差异变化趋势
听力	t=1.87，p=0.062，Cohen's d=0.21	t=2.53，p=0.012，Cohen's d=0.34	t=2.89，p=0.004，Cohen's d=0.39	不显著→显著，差异扩大
语法	t=1.95，p=0.052，Cohen's d=0.23	t=2.76，p=0.006，Cohen's d=0.37	t=3.02，p=0.003，Cohen's d=0.41	边际显著→显著，差异扩大
阅读	t=2.03，p=0.043，Cohen's d=0.24	t=3.15，p=0.002，Cohen's d=0.42	t=3.98，p<0.001，Cohen's d=0.53	显著→极显著，差异大幅扩大
词汇	t=1.79，p=0.074，Cohen's d=0.20	t=2.48，p=0.013，Cohen's d=0.33	t=3.26，p=0.001，Cohen's d=0.44	不显著→显著，差异扩大
写作	t=1.83，p=0.067，Cohen's d=0.21	t=2.51，p=0.012，Cohen's d=0.34	t=2.78，p=0.006，Cohen's d=0.38	不显著→显著，差异扩大

如表 3 可见，T1 数据分析显示仅阅读技能实验组显著高于对照组 (p=0.043)，其余技能两组无显著差异 (p>0.05)，整体基线基本一致，满足干预效果对比条件。T2 数据显示 5 项技能均出现显著组间差异 (p<0.05)，实验组优势初步显现，阅读技能效应量最大 (d=0.42)。T3 数据呈现差异进一步扩大，阅读技能达极显著水平 (p<0.001)，整体均值组间差异极显著 (p<0.001)，效应量达中高水平 (d=0.54)。

综上，通过实验结果可见，评价模型对于教学提升

的效果存在一定滞后性，第二学期才开始呈现一定成效，虽总体成绩对比尚未达到显著效果，但从各项语言技能增值情况来看，实验组教学效果明显，这也达到了设计此模型的初衷。当然，模型的使用和教学实践的具体操作还需要进一步的优化，而且估计还需要为期更长的教学实践从多角度来验证模型的有效性。

5 挑战与对策

首要挑战为教师评价能力缺口问题。实验班老师需要对增值评价 DEA/SGP 模型原理深入理解，并根据

DEA 模型投入、产出指标进行教学设计并准备课堂教学数据,这对老师的教学评价设计能力有一定的挑战:过程性评价要求教师高频记录学生行为,DEA 模型主要反馈高投入低产出预警,对于低投入低产出的学生还需要教师进行学习投入、产出期望值达成率的反馈,实验班教师在教学设计和实施方面需要付出更多精力,可能影响长期实施意愿。

另外,数据质量控制是一大难题。数据收集复杂,投入指标中主观数据和客观数据量化整合也有一定难度。另外,评价指标权重需要个性化调整,教师教学方式个体差异、学生群体差异,评价指标需根据班级具体情况设计。在具体教学实践中,由于需要使用多种语言教学平台,如教学实践期间,实验班老师在进行 DEA 模型数据整合时,需要整合三个教学平台的数据,数据整合难度大。

此外,学生适应性差异也呈现一定难度。高频率量化评价方式可能会造成学生心理压力,出现应付或者抵制的情况。纯 DEA 模型学习效率反馈对低投入低产出的学生无法产生警示反馈,用学习效率雷达图直观反馈学习投入、产出率可能会造成一定学习攀比压力。同时 SGP 模型可能会造成高起点学生可能因“成长百分位提升空间小”而丧失动力。

因此,面对当前挑战,我们尝试建立动态权重调整机制,通过实证研究优化指标体系,简化评价工具操作流程。另外,尝试利用 AI 技术辅助整合多平台数据,并开发数字化评价平台,实现数据自动采集与分析。

6 结论

本研究构建的基于增值评价的综合评价体系模型,通过整合多元评价方式,突破了传统评价的单一性局限,为大学外语教学提供了“关注个体成长、聚焦过程改进”的科学评价工具。本次研究的对象局限于单一专业的部分学生,研究数据有一定片面性,后续研究将扩大实证验证范围,通过进一步的教学实验检验模型的有效性与普适性,并结合人工智能技术优化数据处理流程,进一步提升评价的精准度与效率。

参考文献

- [1] 骆文淑,陈方丁.增值评价中的天花板和地板效应分析[J].中国考试,2024,383(03):87-97.
- [2] 王娟,胡钦晓.美国教师增值评价的实践经验及启示[J].中国考试,2024,382(02):83-91.
- [3] 沈悦,魏佳.英国学校问责制中增值评价的应用及启示[J].郑州师范教育,2024,13(01):28-34.
- [4] 徐璐,钱佳.基于多元证据的美国教师效能增值评价模型及应用启示[J].上海教育科研,2024,449(10):37-45.
- [5] 杨立军,夏紫薇.教育增值评价 50 年:演进、挑战与进路[J].高教发展与评估,2024,40(02):1-18+119.
- [6] 张向众,邓冬琳,乔磊.21 世纪以来我国增值评价研究的回溯与展望——基于CiteSpace的可视化分析[J].曲靖师范学院学报,2024,43(02):94-102.
- [7] 李俊飞,谭顶良.增值评价模型应用的问题探讨与优化策略——以我国 45 项实证研究数据为佐证[J].中国考试,2024,386(06):71-78.
- [8] 王丹.OBE理念下公共英语课程增值性评价体系的构建与实践[J].科教文汇,2024,621(21):128-131.
- [9] 梅晴,严雪娟.基于产出导向法的大学英语课程增值评价教学改革与实践[J].中国多媒体与网络教学学报(中旬刊),2024,(04):38-41.
- [10] 秦盛妮,谭飞刚.增值评价在混合式教学中的应用探索[J].中国教育技术装备,2025,595(01):95-97.
- [11] 柴婷婷,高洁,赵满坤,等.基于增值评价的研究生培养质量评价体系构建思考[J].计算机教育,2024,349(01):56-60.
- [12] 庞红梅.教育数字化转型背景下大学英语增值评价生态体系研究[J].现代英语,2024,(17):77-79.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<http://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS