

计算机辅助二语动态评估系统PELDiaG和CODA的比较研究

何惠勤

西安交通大学, 外国语学院, 陕西 西安
Email: hqhe@mail.xjtu.edu.cn

收稿日期: 2021年1月11日; 录用日期: 2021年2月16日; 发布日期: 2021年2月26日

摘要

本文介绍和比较了两个典型的计算机辅助二语动态评估在线系统PELDiaG和CODA, 从诊断试题命制、干预方式以及迁移任务的设计等方面对比分析二者的异同, 以为智能化的二语动态评估研究的发展提供借鉴。二者从社会文化理论出发, 以动态评估方法为框架, 以计算机和互联网技术为支撑, 用不同的研究方法探索了二语学习者在不同的动态干预模式下的学习行为。PELDiaG系统在诊断试题的构建方面更为严谨, CODA系统则通过设计一种特殊的评分方式来量化学习潜能。PELDiaG系统以多模态的文字图表视频等作为介入手段, 是一种比较显性的介入方式, 而CODA系统对学习者的干预通过逐渐显化干预来实现。

关键词

动态评估, 诊断性评估, 计算机辅助教学, 第二语言评估

Computer-Assisted L2 Dynamic Assessment: A Comparative Study between PELDiaG and CODA

Huiqin He

School of Foreign Studies, Xi'an Jiaotong University, Xi'an Shaanxi
Email: hqhe@mail.xjtu.edu.cn

Received: Jan. 11th, 2021; accepted: Feb. 16th, 2021; published: Feb. 26th, 2021

Abstract

As a learning-oriented and diagnostic assessment approach, Dynamic Assessment (DA) focuses on

the learning process and the learners' individual development. It aims to integrate mediation and instruction into assessment, and thus provides diagnostic information about the learner's weaknesses and strengths, with the hope of pinpointing the source of the students' problems in learning. Recently, DA research has been expanded to the field of second language (L2) education and then computer-assisted language learning setting, as is shown in this study. Two computer-assisted L2 dynamic assessment systems, CODA (Computerized Dynamic Assessment) and PELDiG (Personalized English Learning: Diagnosis and Guidance), constructed respectively by two research teams at the US and Chinese universities, are introduced and compared in a systematic and detailed way. Similarities and differences are discussed with a focus on such DA features as transfer, DA models, and intervention. The study sheds light on the future research on second language assessment as well as L2 computer-assisted teaching and learning.

Keywords

Dynamic Assessment, Diagnostic Assessment, Computer-Assisted Language Learning, Second Language (L2) Assessment

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 动态评估与第二语言学习

1.1. 动态评估(Dynamic Assessment, 缩写为 DA)的理论基础

作为一种以促学为导向的诊断性评估方法(Diagnostic Language Assessment) (Lee, 2015), 动态评估强调教师与学生的互动, 提倡在测量与评估过程中通过运用提示、指导和反馈等干预手段激发学习者的潜能, 使学习者的个体发展达到最大化[1]。评估并非与教学环节相隔离, 而是作为教学环节不可分割的一部分, 与教学活动一起共同致力于发现个体的学习发展特点。作为评估者的教师不再是学习行为的观察者, 而是致力于和学习者共同解决问题、促进学习者进步的中介者(mediator) (Poehner, 2007) [2]。

动态评估的理论基础是前苏联著名心理学家 Vygotsky 的社会文化理论。该理论从社会文化层面探讨学习与发展的相互关系, 提出干预(meditation)、最近发展区(Zone of Proximal Development, ZPD)及支架(scaffolding)等概念(Vygotsky, 1987) [3], 对动态评估的理念与方法有重要影响。其中, 作为核心内容的最近发展区(ZPD)是指学习者在没有外界帮助的情形下的所表现的学习能力和其在相应外界帮助下的所达到的学习能力之间的差异(如图 1 所示)。动态评估的目的正是通过给被评估者提供适当的指导和干预使其学习潜能被进一步激发, 从而达到能力的进步和认知水平的提高(韩宝成, 2009: 452-458) [4]。动态评估是在对传统静态测验的批判和发展中形成的, 它既指一种新的评估理念, 也指该理念指导下的一系列的方法(孔文、方洁, 2013: 57-61) [5]。

在传统静态测验中, 无论是能力水平测验(proficiency test), 分级测验(placement test)或是学业成就测验(achievement test), 测试的主要目的和功能都是对学习者的过去或目前的知识状态提供分数报告。测试设计者和实施者并不重点关心学习者未来的发展, 被试也无法直接从测试中获得反映自身知识状态的诊断反馈信息。动态评估是相对静态评估而言的, 其最大特点是能跨越多个时间点观察评估学生的进步与改变(黄光扬, 2002: 44-48) [6]。与以标准化测试为特征的静态评估相比较, 动态评估的特点是个性化, 过程化, 交互式的。它侧重于在教师(或者具有更高学习能力的人)和学习者的互动中发现学习者的潜能, 它

关心学生学习新知识和新技能的过程(李清华等, 2014) [7]。

动态评估的过程既包括学习者的自主学习过程, 也包括来自外部的干预(Poehner, 2008) [8]。Lantolf & Poehner (2004: 49-74) [9]把动态评估模式大致分为两类: 干预式(interventionist)和互动式(interactionist)。前者是以标准化的逐步文本提示为特征的, 有次序的和循序渐进的干预方式, 而后者则是以中介者(如教师)与学习者之间的对话为干预手段的评估模式。

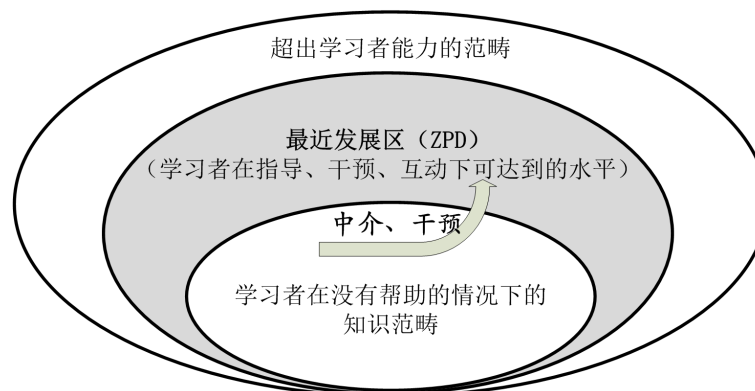


Figure 1. Illustration of ZPD and Mediation
图 1. 最近发展区(ZPD)和中介示意图

1.2. 动态评估在二语教育领域的发展

近年来, 动态评估在教育评估中受到越来越多的关注。国际教育评估期刊 *Assessment in Education: Principles, Policy & Practice* 在 2011 年第二期出版了关于动态评估研究的专刊 “Addressing issues of access and fairness in education through Dynamic Assessment”。把动态评估引入到二语教育和评估领域, 并对二语动态评估做出积极尝试, 包括美国宾州州立大学 James Lantolf 和 Matthew Poehner 的研究团队。Lantolf 教授所带领的团队长期从事二语(L2)动态评估方面的研究, 并出版了由 Poehner 撰写的首部关于二语动态评估的著作 *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*, 该书作者把动态评估的应用拓展到二语发展领域, 是在应用语言学范围内做出的首创性贡献(侯杰、刘燕, 2011: 89-92) [10]。在动态评估和二语教学框架中, 有一些实践者做出了积极地探索和尝试。例如 Ableeva & Lantolf (2011)对以法语为二语的听力学习者进行了研究, 发现学习者在动态干预下对听力素材中的内容回顾能力(recalling ideas)有提高, 并且在一定程度上可以将这种能力运用到难度更高的任务中[11]。

动态评估理论涉及的领域与相关内容也引起国内应用语言学者的兴趣, 韩宝成(2009: 452-458) [4]以及孔文和方洁(2013) [12]对动态评估在外语教学中的运用做了系统的综述。一些学者尝试把动态评估研究同二语教学结合起来, 也取得了阶段性的成果, 如彭定金(2004: 26-31) [13]为大学英语课堂教学设计了一套“五段式课堂教学动态评估”流程和四种评估策略。张艳红(2008: 73-81, 2010: 46-50) [14] [15], 将动态评估引入到网络写作教学, 以网络技术(如电子邮件)为支持对学生的写作教学进行干预。孔文等(2013: 77-86) [5]进行了 L2 写作动态评估中同伴中介干预和教师中介干预的比较研究。孟亚茹等(2019) [16]探索了基于听力认知属性的动态评估干预模式有效性。李奕华(2015: 59-67) [17]探讨了在动态评估理论框架之下, 教师反馈、批改网自动评分以及“教师干预+批改网评分”三种反馈方式对大学英语专业本科生写作教学的影响, 发现由教师参与的反馈方式更有利于提高写作水平。

动态评估由于对学生和评估者之间的互动有较高要求, 因此已有的研究多在传统课堂教学背景中。

Poehner 和 Lantolf 在 2008-2011 年进行了计算机化的法语、俄语及汉语语言水平动态评估研究, 利用所设计开发的“计算机化动态评估(Computerized Dynamic Assessment)系统 CODA”对动态评估的迁移任务和学习潜能进行了探索性的研究(Poehner & Lantolf, 2013: 323-342) [18]。在国内, 包括笔者在内的科研团队搭建了“个性化英语学习诊断系统 PELDiaG” (Personalized English Learning: Diagnosis and Guidance)。它是依托计算机网络技术对英语听力和阅读提供个性化学习指导以及进行动态诊断和评估的数字化平台。我们将在后面对这两个计算机辅助二语动态评估系统做详细的介绍和对比。

2. 动态评估的核心概念

2.1. 中介(mediation)

DA 中的中介(mediation)指对学习者的学习过程进行的介入或干预, 是实现过程性、交互性评估的具体手段和操作模式。在 L2 动态评估中, 中介的内容既包括所需要的语言知识(词汇, 语法等), 也包括语言学习策略(认知和元认知策略)。担任中介角色的中介者可以是评估者、教师或者是有更高学习能力的人, 他们在评估的过程中同时也是被试的合作者, 观察者和鼓励者。事实上, 正是由于中介的不同实现途径, 不同表现形式以及不同的目的, DA 的模式呈现出不同的特点, 主要包括介入式动态评估和互动式动态评估。

介入式动态评估(Interventionist DA), 以 Budoff 的“学习潜能测量模式”(Learning Potential Measurement Approach)、Guthke 的“学习测验模式”(Lerntest Approach 或 Leipzig Learning Test)、Carlson and Wiedl 的“极限探测评估模式”(Testing-the-Limits Approach)以及 Brown & Ferrara 的“渐进提示模式”(Graduated Prompt Approach, GPA)为代表, 强调中介的标准化过程, 适用于学习者人数较多的情况。在这样的介入模式中, 提示、暗示或是有导向性的提问都按照严格的文本逐步提示来操作, 从隐性到显性, 从抽象到具体, 使评估的客观性最大化。虽然介入式 DA 同 Vygotsky 的介入概念有所不同, 但这种模式的优点是可以量化分数结果, 使得研究者可以用传统静态测验的构念来讨论学习者的动态评估表现。正如 Poehner & Lantolf (2013: 323-342) [18]指出的, 虽然标准化的介入方式限制了干预者的灵活性, 这种“万能”(one-size-fits-all)的提示模式和干预方式仍有可取之处: 它可以在同一时间高效地评估更多的学习者。

互动式动态评估(Interactionist DA), 以 Feuerstein 等(1988) [19]的中介学习经验理论(Mediated Learning Experience, 简称 MLE)为代表, 则是指以学习者和中介者(教师)之间对话为特征的 DA 范式。该模式把教学和评估融为一体, 强调教育者、学习者和学习任务三者之间不断相互作用。Feuerstein [20]等提出的包括“学习潜能评估模型”在内的一些动态评估应用模式, 强调中介的非标准化、个体交流以及高度交互。中介者在测试过程中根据学习者的不同表现灵活地对任务作出随时调整, 并关注学习过程的情感因素, 指导他们解决问题的技巧、方法和思考过程, 最终让学习者具有独立完成问题或解决问题的能力(范兆兰, 2006) [21]。Feuerstein 的模式“反映了 Vygotsky 的 ZPD 思想, 给测验者较大的自由与学习者互动, 通过一系列有针对性的帮助使学习者的潜能得到发展(韩宝成, 2009)” [4]。与此同时, 与介入式 DA 相比, 互动式 DA 需要投入大量的时间、金钱和人力, 并且对评估介入者的专业知识和能力有较高的要求。

2.2. 迁移(Transfer or transcendence)

迁移, 即 transfer 或 transcendence (Poehner, 2007: 323-340) [2], 是指学习者在掌握了某种学习技能或认知概念后把它运用到任务类型相似(但不相同)、任务难度相同或更高的新任务中的能力。根据新任务的难度差别, 可分为近迁移(任务难度变化不大), 远迁移(任务难度变化较大), 和极迁移(任务难度大幅变化) (Brown & Ferrara, 1985) [22]。

Poehner (2007: 323-340) [2]对迁移任务的研究表明,在完成熟悉任务中表现近似的学习者在把这种能力运用到解决新问题过程中却表现出很大的差别。因此在评估中包含类型相似而难度更大的任务能够使我们可以更好地了解学习者是否内化了之前获得的中介内容。迁移任务也是“渐进提示模式”对 DA 的主要贡献之一。迁移任务的设计可以看作是动态评估的一部分,与动态评估“前测—介入—后测”过程融合在一起。

3. 计算机辅助二语动态评估: CODA 和 PELDiaG

Poehner & Lantolf 利用所设计的“计算机化动态评估系统 CODA”对法语、俄语及汉语进行了 L2 阅读和听力动态评估研究(Poehner & Lantolf, 2013: 323-342; Poehner, Zhang & Lu, 2015: 337-357) [18] [23], 试图发现学习者在计算机干预时和无干预时的所表现出的不同学习能力, 预测学习潜能, 揭示“干预提示”融合在测试中的学习证据。“个性化英语学习诊断指导系统 PELDiaG”是以心理测量和二语习得理论为基础, 从动态评估和认知诊断的角度出发, 所构建的诊断 L2 听力和阅读学习技能的测试模型(马晓梅等, 2012: 59-63) [24]。该系统的构建是基于动态评估理论之上的, 在给予学习技能诊断的前提下, 通过计算机技术手段给学习者提供及时的和个性化的反馈和干预, 将诊断、评估和网络教学相结合, 让学习者在测试和评估的环节中得到“学习和反馈”, 在“测试评估—反馈干预—测试评估”的动态交互模式中促进其发挥学习潜能。

3.1. PELDiaG 动态评估系统的建立

“个性化英语学习诊断指导系统 PELDiaG”是面向大学英语非专业一年级学生的英语视听和阅读学习动态诊断和指导平台。平台由能力诊断、策略指导以及强化训练三个模块构成。学生通过完成在线诊断测试题, 可立即获得包括水平分数和技能评估在内的诊断报告。诊断报告同时根据学生在视听和阅读不同认知技能方面所表现的特点通过热链接形式给出进一步策略指导及强化训练。

PELDiaG 系统的干预模式构建包括基于诊断试题的动态诊断反馈、视听及阅读技能指导以及技能强化训练。

3.1.1. PELDiaG 诊断试题构建

我们首先构建了用于进行诊断测试的试题。以视听测试为例, 该系统测试题依据 Anderson (1995) [25] 的认知理解过程三段式: 感知—切分—应用理论以及 Richards (1983: 219-240) [26] 对于听力过程中所涉及的微技能分类, 参照《简明英语测试教程》(邹申、杨任明, 2000: 116-117) [27] 中对听力技能的分类, 根据《大学英语课程教学要求》以及大学英语四级考试大纲中对听力理解的要求, 构建了适合大学一年级非英语专业学生视听评估诊断测试模型, 共有评估诊断试题三套(3 Tests), 每套试题包括短对话、长对话、篇章、视频和听写共 24 个测试题目。在和测试题目相匹配的前提下, 每套测试题包含 7 个 L2 技能(即属性¹, 标记为 A 加数字): 辨音能力(A1)、词汇知识(A2)、句子结构和语法(A3)、推理能力和文化背景知识(A4)、捕捉细节能力(A5)、主旨能力(A6)、记笔记和听写能力(A7)。每个属性对应于不少于三个测试题(如表 1 所示)。

试题所测试的认知技能/属性确定步骤如下。首先, 由学科专家(英语教师)对试题进行属性标注: 共由 9 人完成, 包括 6 个非命题专家(长期从事 L2 教学的教师, 并且接受了属性定义培训), 3 个命题专家(试题构建者)。每个题目根据相关的技能/属性按主次进行标注。此外, 对 17 个 L2 (非英语专业和英语专业) 学生进行有声思维培训和数据收集, 用于辅助判断专家标注的结果。最后根据属性标注结果生成 L2 学习者视听认知属性权重评估表。如 Test 1 第 17 题的属性权重分布为: A5 属性 3 分、A4 属性 2 分、A7 属

¹这里借鉴了认知诊断学科术语——属性(attribute), 来表示认知技能, 即测试题目所涉及到的认知过程、策略、技能, 以及知识元素。

性 1 分。这样,学生在完成试题(Test 1)的测试后,系统就能按照试题中属性权重分数给出学习者认知属性特征的判断。如某个属性得分比较高,则表明学生对该属性的掌握程度比较好,反之则表明对该认知属性的掌握有欠缺。此外,系统同时也参照大学英语四级考试的得分标准,赋予题目水平测试权重分数,在测试结束后给出直观的水平测试分数。系统中的阅读评估测试模型也按照类似的步骤完成,由于篇幅所限,就不在此赘述。

Table 1. Definitions of the attributes in audio-listening assessment

表 1. 视听评估测试模型中的属性定义

序号	属性编号	属性(技能)解释	对应题目序号 (属性假设)
1	A1	辨音能力:能通过语音知识如辨音、连音、语调等语音知识识别信息	题目: 1, 2, 3
2	A2	词汇辨理解:听力中低频的单词,口语常用词组和俚语表达的辨认和理解	题目: 4, 5, 6
3	A3	句子结构和语法功能能力:影响句义的句法结构如虚拟倒装否定强调等	题目: 7, 9, 10
4	A4	事实和细节抓取能力:时间、地点、人物关系等	题目: 8, 11, 15, 18, 21
5	A5	主旨能力:准确认知、归纳材料所传达主要信息	题目: 14, 17, 20
6	A6	通过上下文语境、背景知识和相关文化常识推测说话者动机、目的、原因交际功能和隐含意义	题目: 12, 13, 16, 19
7	A7	短时记忆和记笔记:通过笔记和短时记忆抓住信息完成相应任务	题目: 22, 23, 24

3.1.2. PELDiaG 干预模式构建

动态评估中最重要的一环是中介/干预,也就是教师或评估者对学生的干预、指导和帮助。在 PELDiaG 系统中,根据认知属性和水平能力评估的结果,对学生的指导和干预从以下几个方面介入(如图 2 所示):

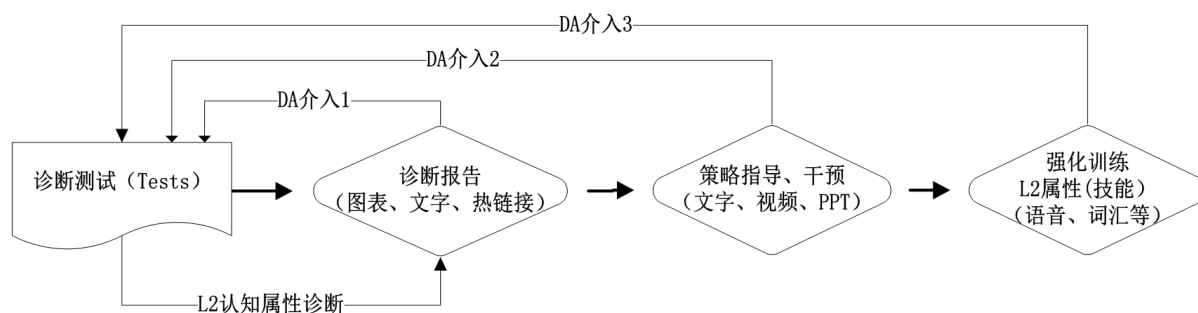


Figure 2. Dynamic assessment and mediation framework in PELDiaG

图 2. PELDiaG 系统中动态介入和评估模式

i) PELDiaG 诊断评估报告

PELDiaG 系统为学习者提供的不是简单的分数结果,而是包含认知属性判断的立体综合分析报告,提示学习者的 L2 认知能力特点,并结合水平测试结果给出相应的建议。综合分析报告以文字、图表和热链接的方式呈现。学习者查看水平分数结果的同时,可以看到正确答案详情、属性评估结果,以及基于以上判断为学习者提供的热链接以使学习者对评估报告内容有进一步的了解。PELDiaG 系统中的诊断报告作为一种 DA 介入方式,在很大程度上克服了传统测试只报告分数,而无法让学习者从测试中获得更多关于学习技能相关信息的弊端。它为学习者提供了初步的动态介入,使得学习者能够了解自己在 L2 视听和阅读学习方面的强势和弱势。

ii) PELDiaG 策略指导

动态评估与传统静态测试的一个重要区别是动态评估强调“评估与教学相结合”(范兆兰, 2006; Poehner, 2007: 323-340) [2] [21], 以及如何帮助学习者克服障碍。评估不是同教学相对立的, 而是作为教学不可分割的一部分, 评估、指导和其他教学活动相结合共同帮助和理解学习者的个体认知发展。因此作为干预手段之一的指导环节对动态评估的最终效果起着至关重要的作用。基于网络的 PELDiaG 系统利用现代教育技术的优势为学习者提供了以下 4 种形式的指导内容, 包括 1) 按照 L2 认知属性分布并由一线教师结合实际教学经验和二语习得领域的研究成果撰写的二语视听和阅读学习文字策略指导; 2) 采访国外二语习得领域知名专家, 如 Rebecca Oxford, James Lantolf, Rod Ellis 等所编辑的视频策略指导; 3) 对成功学习者和不成功学习者的采访视频策略指导, 4) 以 PPT 动画呈现的学习策略指导。PELDiaG 系统的指导干预同时也是基于多元智能和多模态理论的介入模式, 为学习者提供了充分的个性化教学和辅导。

iii) PELDiaG 视听/阅读技能强化训练

PELDiaG 系统对学习者的第三种干预方式是技能强化训练部分。动态评估的最终目标是希望学习者借助评估和干预这一过程, 获得更高的学习能力或达到更高的学习水平, 这种能力在缺乏外界帮助下是学习者本身所无法企及的。考察学习者是否已经“内化”了所获得的更高能力, 可以通过赋予学习者相同或者类似的任务, 来考察这样的学习技能是否已经潜移默化地被学习者掌握。这种完成相同或者类似任务的能力被称之为迁移能力。同时, 通过完成这类任务还可以让学习者进一步巩固和强化已掌握的技能。因此, 迁移任务的设计对动态评估过程在很多情况下不可或缺, 它在考察动态评估教学效果的同时, 对促进学习者个体进一步发展从而接近最近发展区(ZPD)也具有非常的意义。

PELDiaG 系统中对迁移能力的考察和巩固是通过强化训练模块来实现的。在这个模块中, 阅读和听力任务首先按照从易到难分为 Level A, Level B, 以及 Level C 三个等级。学习者通过完成不同等级的试题可以更真切地了解自己的学习状态。其次, 在听力和阅读的每个等级中, 试题的分布是按照属性编排的, 即视听强化分为: 语音语调(辨音、连音和语调), 词汇、语法和结构, 细节, 推理和文化背景(上下文推理和文化背景知识), 长对话和短文(主旨、细节、推理), 听写(记忆和记笔记)以及视频(主旨、细节); 阅读强化包括: 猜测词义, 句子理解, 句子推断, 段落理解, 段落推测, 主旨概括, 细节捕捉以及阅读速度训练。强化部分的题目虽然也是按照动态评估部分的属性进行编制, 但是考虑到题目素材本身归类的便捷问题, 有些题目在按照属性排列的同时, 也按照题目本身的特点进行归类, 比如长对话和短文中包含了主旨、细节和推理三个技能, 而这三个技能在别的试题中也有体现。

3.2. CODA 的动态评估模式

CODA 系统的特点是运用了渐进提示模式(如图 3 所示)。如果答题者第一次尝试就答对题目, 则得到该题满分 4 分; 如果第一次答错, 则获得第一次提示(在原文中缩小正确答案的范围), 在此基础上如果答对则得 3 分, 如果打错则继续得到提示直到答对为止, 并且据此得到相应的 2 分, 1 分和 0 分。Poehner & Lantolf (2013) [18]称这个评估过程为“单回合”(microgenesis), 即干预过程是即时性的(moment-to-moment)。CODA 的题目中第二语言为汉语和法语, 题目分别为汉语听力及阅读以及法语阅读。其中听力是短文理解, 在两遍短文播放结束后给出问题。汉语阅读题目仍然是短文理解, 而法语阅读题目是长篇阅读理解。问题的形式是标准化测试的多项选择题, 为了最大化可提示的次数, 每道题目有 5 个选项。问题和 5 个选项都用英语给出, 以排除因为看不懂题目而造成对原文理解能力评估的偏差。阅读题目的构成包括: 词汇, 语法, 语篇以及文化。听力测试的构念(construct)同阅读类似, 但是多了对语音的考察²。

²在另一篇文章中, Poehner, Zhang & Lu (2015: 337-357)分别把 CODA 系统中听力和阅读的试题构念做了新的界定。其中听力为: 语音, 语法(时态), 语用, 上下文推理以及信息捕获; 阅读为: 词汇, 结构, 语篇和体裁。

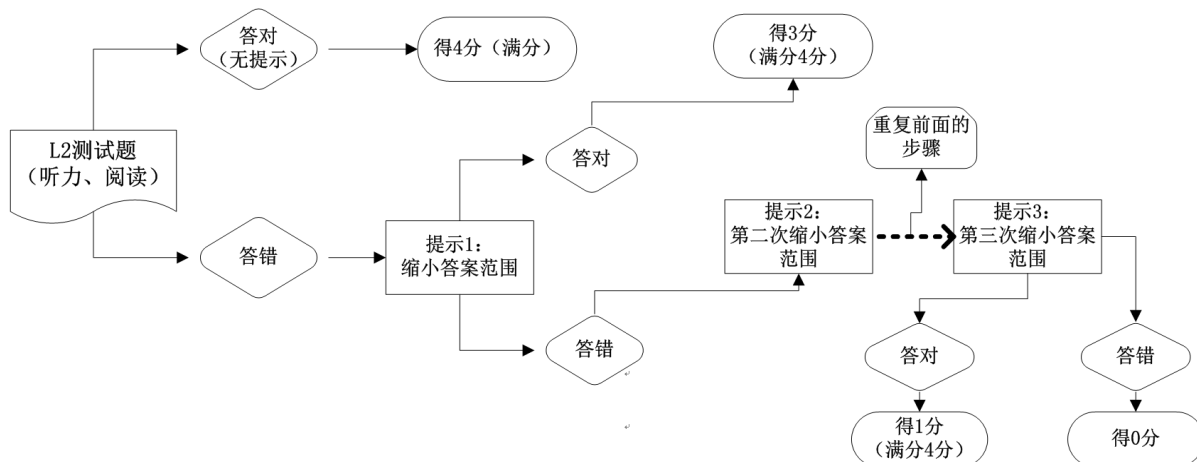


Figure 3. Computerized dynamic assessment in CODA
图 3. CODA 系统的动态评估过程示意图

除了利用计算机技术对测试者进行标准化的文本逐步提示，CODA 系统还借鉴了学习潜能得分这一概念试图量化动态评估的效果。学习潜能得分(Learning Potential Score, LPS)是由 Sternberg and Grigorenko (2002) [28]提出的用于量化学习者在干预前和干预之后的表现差异。基于社会文化理论的动态评估关注学习的过程，以及在外部的帮助和介入的情况下学习者能够取得的最大成就。学习潜能得分这一概念的提出迎合了这一理念，借助公式计算来量化学习者在获得干预之后可能产生的水平提高。学习潜能得分公式如下：

$$LPS = \frac{(S_{post} - S_{pre})}{S_{max}} + \frac{S_{post}}{S_{max}} = \frac{2S_{post} - S_{pre}}{S_{max}}$$

其中 S_{max} 表示可能获得的最高分数， S_{pre} 是学习者在没有获得干预情况下的分数，即假设学习者只有一次答题机会而获得的分数。 S_{post} 表示干预后的分数，也就是学习者在得到提示后正确作答的分数。按照前面的原则， S_{post} 不可能是满分 4 分，而是相应地根据被提示的次数而只能是 3，2 或者 1 分。因此个体的 LPS 是在学习者自身内部之间所进行的比较，它反映了学习者初始状态和最终能够达到的能力之间的差异。LPS 越大表示学习者相比过去所产生的进步越大，也表示干预和指导所产生的作用可能越大。Poehner & Lantolf (2013: 323-342) [18]借鉴了学习潜能得分的计算方式，并对不同学习者的个体表现进行了讨论。

4. 讨论：CODA 和 PELDiaG 的对比

结合前面对两个系统的分别介绍，不难看出这两个计算机化的动态评估(C-DA)在某些方面有共同之处。比如，无论是 CODA 还是 PELDiaG 都同课堂动态评估有所不同。它们都运用了计算机技术参与到测试和介入的过程，并在此基础之上对数据进行收集。这样的评估方式同传统课堂“一对一”模式的动态评估相比，克服了其效率低，无法造福大规模学习者的弊端。但 C-DA 也有自身的局限性，即这种计算机程式化的介入和干预方式导致 DA 介入的灵活性大大下降。作为中介者的计算机无法面对面地观察学习者的现场表现，而 DA 提示的可能性受到这种模式的限制，无法最大化介入干预的可能性。此外，这两个测试评估系统都属于无风险测试，即测试评估的结果并不对某些重要事件(如升学，分级或者就业)产生影响。

然而在目前为数不多的计算机化动态评估模式中，CODA 和 PELDiaG 也呈现出各自的不同特点(见表 2)。

Table 2. A comprehensive contrastive analysis between PELDiaG and CODA

表 2. PELDiaG 和 CODA 系统动态评估模式对比

DA 评估过程	C-DA 系统	PELDiaG	CODA
评估的类型		动态评估; 诊断性评估; 形成性评估	动态评估; 诊断性评估; 形成性评估
设计试题的目的		提供显性的诊断性评估(诊断测试); 提供补救教学和训练(强化训练试题)	区分学习者在干预下的表现, 预测学习潜能, 揭示“干预提示”融合在测试中的学习证据
C-DA 试题命制		L2 教师根据学习技能编制试题; 学科专家参与试题属性标定 题目数量: 24 (items) × 3 (tests) = 72	以汉语听力为例: 汉语水平考(HSK)试练习题; 增加一个选项使之达到 5 个; 题目数量: 23
C-DA 项目属性/试题构念		项目属性(马晓梅等, 2012) [23] 听力: 辨音能力、词汇知识、句子结构和语法、推理能力和文化背景知识、细节捕捉、主旨能力、记笔记和听写能力 阅读: 猜测词义、句子理解、句子推断、段落理解、段落推测、主旨概括、细节捕捉、阅读速度	试题构念(Poehner, Zhang & Lu, 2015) [22] 听力: 语音, 语法(时态), 语用, 上下文推理, 信息捕获 阅读: 词汇, 结构, 语篇, 体裁
动态干预的方式(介入方式和内容)		互动式动态评估; 显性 介入内容: 二语学习能力和技能的诊断性报告(文字); 多模态形式的策略指导和干预; 按照二语技能编制的强化干预题库	介入式动态评估; 隐性 介入内容: 答题过程中给予在线文字提示(通过缩小阅读或听力范围来进行提示, 每一次提示都让答题者更接近正确答案所在的文本)
动态干预的方式(介入过程)		1.多回合评估(multi-session assessment) 2.测试完毕, 给予两种形式的反馈和干预: 多媒体诊断报告(图表, 文字, 热链接), 多媒体学习策略指导干预(文字、动画、视频(专家访谈、成功学习者案例)) 3.不同学习者的诊断报告不同	1.单回合评估(microgenesis) 2.对错误答案进行逐步提示: 最大值为三个回合 3.one-size-fit-all 学习者获得的提示内容不变
干预目的		帮助学习者诊断英语听力及阅读学习强项和弱项, 提供实时反馈和指导, 培养学习策略意识	以计算机为手段, 观察二语学习者对动态评估中的干预反应, 考察计算机逐步干预的学习效果
迁移任务(Transfer Items)的设置		在强化训练模块中, 按照阅读及听力属性专门编排反映迁移能力的试题, 按照难度从低到高	在试题本身靠后的地方, 用以考察学习者是否掌握了从动态评估中获取的新技能; 内嵌于测试当中; 量化考察
对学习者的训练(强化)		按照属性/技能编制的不同级别的强化试题	无
学习潜能得分的计算		无	有
测试风险		无风险(用于学习者自测或自主学习)	无风险
目标学习者(用户)		开放性学习平台; 在校本科非英语专业一年级大学生, 500 名以上用户。测试题目是英语, 学习者母语是汉语	研究平台; 样本量为 163 名答题者, 测试题目为汉语和法语, 学习者母语为英语

首先, 在相关文献中(Poehner & Lantolf, 2013; Poehner et al., 2015) [18] [23], CODA 系统用于进行动态评估的试题没有进行专家标定, 使得试题的构念不够严谨。作为一种诊断性评估, 动态评估中试题的构念至关重要。也就是说, 每个题目是否测试了相关的语言技能这一指标应得到充分验证。在 PELDiaG 中, 从语言技能/属性的定义到试题属性的标定, 及诊断结果的报告, 都比较严谨地论证了试题的可靠性。

其次, 二者介入方式和目的不同。这也是两个 C-DA 系统在用户界面最明显的差别所在。在 CODA 中, 介入是按照逐步提示法进行, 通过逐步缩小正确答案范围对学习者的动态干预, 并进一步根据被提示的次数来统计学习者潜能得分的。这种介入方式被 Poehner & Lantolf (2013) [18]称之为“单回合”动态评估, 即动态介入是在诊断测试中的一个题目(item)中完成的。同时, 这样的干预方式也是一种介入式动态评估, 即干预是按照逐步的文本提示进行的。在 PELDiaG 中, 计算机对学习者的动态干预是在一个诊断测试(test)结束之后进行的, 体现在文字和图表诊断结果报告和以热链接为形式的进一步学习技能指

导中。学习者在整个系统中可以被多次评估,是一种“多回合”的动态评估方式(multi-session assessment),同时也是以每个学习者得到不同的诊断反馈为特征的互动式动态评估。与传统语言测试(非诊断性语言测试以及静态测试)相区别的是,PELDiaG系统给予学习者在相关学习技能方面详细的文字报告,并对如何进行下一步学习提出补救性建议。

此外,两个平台的构建和干预目的也不完全相同。PELDiaG系统注重第二语言的学习功能,通过帮助学习者诊断英语听力及阅读学习强项和弱项,提供实时反馈(文字诊断报告)和指导(学习策略链接),培养学习策略和意识。系统中除了诊断功能以外,还包含了大量的语言学习辅助资料,如英语听力和阅读学习策略指导,专家访谈视频,成功学习者和不成功学习者的案例对比等,是一种“显性”的干预方式,学习者可以通过“评估-指导-再评估”的过程有意识地培养自己在学习技能方面的提高。CODA系统更接近一个小规模的实验平台,它试图通过观察二语学习者对计算机化动态评估中的干预反应,考察计算机逐步干预之下的学习效果,计算机对学习者的干预是一种比较“隐性”的干预方式。两种介入方式都被发现是有用的。在CODA中,三组阅读和听力的学习者成绩在干预之下都有不同程度的提高。在PELDiaG中,以视听学习为例,在对255个学习者的网络问卷调查中(见表3),86.8%的学习者认为该学习干预模式总体对他们的学习能力发展有促进作用(“很有帮助”或者“较有帮助”)。

Table 3. Survey results on PELDiaG learning
表 3. 网络问卷调查: PELDiaG 对视听能力发展的促进作用

满意程度	视听行为诊断%	策略指导%	总体
很有帮助	21.4	24.8	29.7
较有帮助	56.4	56.4	57.1
一般	17.7	15.8	11.3
没多大帮助	4.1	3	1.9
没有帮助	0.4	0	0

在迁移任务的安排上,二者也有各自的特点。CODA系统为了让学习潜能可以被量化,将包含迁移任务的试题(transfer items)内置于整个测试之中。学习者在完成前半部分试题后被假设为应该已经获得了相应的语言学习技能,并且能够运用这些技能解决类似的任务。Transfer items的设置是为了量化考察这个假设是否成立。Poehner & Lantolf (2013) [18]的研究显示, CODA的计算机化动态评估模式在观察学习行为方面是有效的。根据学习潜能得分可以看出,原始分数相同的学习者在获得同样的提示和干预之后却表现出不同的学习能力(干预分数不同),这表明具有相同学习水平的人对同样教学的吸收能力可能并不相同,也就是说他们拥有不同的“学习潜力”(Poehner, Zhang & Lu, 2015) [23]。这样的信息能够为教育者在未来教学中提供更有针对性的决策依据。PELDiaG系统中没有内置迁移试题或任务,而是在独立于动态诊断试题之外构建了训练试题部分,并根据题目的难度程度不同由易到难分成 Level A, Level B 和 Level C 三个级别。这些题目和前面的动态诊断试题相呼应,按照二语听力和阅读学习技能编排,使得学习者在获得诊断评估之后能够进一步强化巩固所学技能。可以看出, CODA通过设计内置迁移任务来测量学习潜能; PELDiaG注重通过让学习者完成迁移任务来内化所学知识,帮助个体接近最近发展区(ZPD)。

5. 结语

《大学英语教学指南》(2020版)指出要重视形成性评价,把形成性评价作为大学英语评价与测试的一项重要任务(金艳、孙杭, 2020) [29]。在当今高等教育国际化的背景下,教学评价体系必须改变以考试为

导向的量化评估标准,并提倡在教学评估中采用诊断性评价方式(蔡基刚,2013) [30]。随着计算机辅助教学在我国外语教学领域的大面积普及,探索网络和计算机等现代教育技术对教育测试与评估的影响势在必行,但目前这些研究大多是基于总结性评估模式和标准化测试(李清华等,2014) [23]。因此如何在把现代教育技术同形成性评估、诊断性评估结合起来,使其促学潜力在计算机网络教学中得到相应的发展和运用,是非常有意义的研究目标。

计算机辅助动态评估以促学为导向,以诊断和干预为手段,以计算机网络技术支持,是对二语动态评估和新的语言网络教学模式的发展所做的有益性探索。它不只是对评价方法的改革,也是一种教育理念和教学方法的革新。动态评估作为形成性评估大家庭中的重要成员,强调把评估同教学相结合,为学习者提供及时的、诊断性的评估反馈。它关注学习者的发展,克服了传统及网络评估测试中“只评估,不诊断,不干预指导,不补救”的弊端。本文介绍和比较了两个计算机辅助动态评估系统 PELDiaG 和 CODA 各自的特点,其中特别介绍了 PELDiaG 系统的构建过程,包括诊断试题的命制,反馈及干预方式以及迁移任务的设计。二者从社会文化理论出发,以动态评估方法为框架,以计算机和互联网技术为支撑,用不同的研究方法探索了二语学习者在不同的动态干预模式下的学习行为,具有各自的不同特点。PELDiaG 系统在诊断试题的构建和验证方面更为严谨, CODA 系统则通过设计一种特殊的评分方式来量化学习潜能。PELDiaG 系统以多模态的文字图表视频等作为介入手段,是一种比较显性的介入方式,学习者通过“测试-诊断反馈(学习指导)-再测试”的模式获得学习进步。平台的学习功能比较突出,是开放式的 L2 英语学习平台。CODA 系统对学习者的干预相对而言比较隐性,通过不断缩小正确答案所在的范围来实现,它关注平台的研究功能,只开放给相关研究人员及受试者。

在两个系统中,中介对学习者的干预灵活性都比传统课堂低,用计算机和网络技术模拟面对面的授课和动态评估过程无论在目前还是未来一段时间内都是有挑战性的课题。但是正如 Poehner & Lantolf (2013) [7]所指出的,就同一时间所评估的学生数目而言,C-DA 的优点是提高了中介效率,从而使其具备了大规模使用的条件。基于计算机和网络技术的 C-DA 所具有的优点如反馈迅速及时、多媒体呈现方式、大容量数据存储、信息反复存取等为动态评估注入了新的活力,为第二语言的评估和教学提供了新的思路,也对目前我国高校迅速发展的大学英语网络教学的模式构建有一定的借鉴意义。

基金项目

国家社科基金项目“构架从认知诊断到动态干预的英语听力模型研究”(16BYY096); 国家社科基金项目“英语写作认知诊断测评模式构建及有效性论证”(17BYY015)。

参考文献

- [1] Lee, Y. (2015) Diagnosing Diagnostic Language Assessment. *Language Testing*, 32, 299-316. <https://doi.org/10.1177/0265532214565387>
- [2] Poehner, M.E. (2007) Beyond the Test: L2 Dynamic Assessment and the Transcendence of Mediated Learning. *The Modern Language Journal*, 91, 323-340. <https://doi.org/10.1111/j.1540-4781.2007.00583.x>
- [3] Vygotsky, L.S. (1987) Thinking and Speech. In: Rieber, R.W. and Carton, A.S., Eds., *The Collected Works of L.S. Vygotsky: Volume 1: Problems of General Psychology*, Plenum, New York, 39-285.
- [4] 韩宝成. 动态评价理论、模式及其在外语教育中的应用[J]. 外语教学与研究, 2009, 41(6): 452-458.
- [5] 孔文, 李敦东, 余国兴. L2 写作动态评估中同伴中介干预和教师中介干预比较研究[J]. 外语界, 2013(3): 77-86.
- [6] 黄光扬. 动态评价的理论基础、主要模式及可用性初探[J]. 比较教育研究, 2002, 23(11): 44-48.
- [7] 李清华, 王伟强, 张放. 形成性评估研究[M]. 北京: 科学出版社, 2014.
- [8] Poehner, M.E. (2008) *Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting Second Language Development*. Springer Publishing, Berlin.

- [9] Lantolf, J.P. and Poehner, M.E. (2004) Dynamic Assessment: Bringing the Past into the Future. *Journal of Applied Linguistics*, **1**, 49-72.
- [10] 侯杰, 刘燕. 《动态评估: 从维果斯基理论角度理解和促进二语发展》评介[J]. 外语界, 2011(4): 89-92.
- [11] Ableeva, R. and Lantolf, J.P. (2011) Mediated Dialogue and the Microgenesis of Second Language Listening Comprehension. *Assessment in Education: Principles, Policy & Practice*, **18**, 133-149.
<https://doi.org/10.1080/0969594X.2011.555330>
- [12] 孔文, 方洁. 另类评估的新发展: 动态评估面面观[J]. 解放军外国语学院学报, 2013, 36(1): 57-61.
- [13] 彭金定. 大学英语课堂教学动态评估[J]. 外语界, 2004(3): 26-31.
- [14] 张艳红. 大学英语网络写作教学的动态评估模式研究[J]. 外语界, 2008(4): 73-81.
- [15] 张艳红. 大学英语写作教学的动态评价体系建构[J]. 解放军外国语学院学报, 2010(1), 33(1): 46-50+127-128.
- [16] 孟亚茹, 刘丹, 晏艺赫, 等. 基于听力认知属性的动态评估干预模式初探[J]. 语言教育, 2019(4): 43-50.
- [17] 李奕华. 基于动态评估理论的英语写作反馈方式比较研究[J]. 外语界, 2015(3): 59-67.
- [18] Poehner, M.E. and Lantolf, J.P. (2013) Bringing the ZPD into the Equation: Capturing L2 Development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, **17**, 323-342.
<https://doi.org/10.1177/1362168813482935>
- [19] Feuerstein, R., Rand, Y. and Rynders, J.E. (1988) Don't Accept Me as I Am: Helping Retarded Performers Excel. Plenum, New York. <https://doi.org/10.1007/978-1-4899-6128-0>
- [20] Feuerstein, R., Rand, Y. and Hoffman, M.B. (1979) The Dynamic Assessment of Retarded Performers: The Learning Potential Assessment Device, Theory, Instruments, and Techniques. University Park Press, Baltimore.
- [21] 范兆兰. 动态评估理论与应用研究[D]: [博士学位论文]. 南京: 南京师范大学, 2006.
- [22] Brown, A. and Ferrara, R.A. (1985) Diagnosing Zones of Proximal Development. In: Wertsch, Ed., *Culture, Communication and Cognition: Vygotskian Perspectives*, Cambridge University Press, Cambridge, 273-305.
- [23] Poehner, M.E., Zhang, J. and Lu, X. (2015) Computerized Dynamic Assessment (C-DA): Diagnosing L2 Development According to Learner Responsiveness to Mediation. *Language Testing*, **32**, 337-357.
<https://doi.org/10.1177/0265532214560390>
- [24] 马晓梅, 孟亚茹, 何惠勤, 刘睿. 个性化英语视听诊断指导模式构建及其系统开发[J]. 外语教学, 2012(5): 59-63.
- [25] Anderson, J.R. (1995) *Cognitive Psychology and Its Implications*. Worth Publishers, New York.
- [26] Richards, J.C. (1983) Listening Comprehension: Approach, Design, Procedure. *TESOL Quarterly*, **17**, 219-239.
<https://doi.org/10.2307/3586651>
- [27] 邹申, 杨任明. 简明英语测试教程[M]. 北京: 高等教育出版社, 2000.
- [28] Sternberg, R.J. and Grigorenko, E.L. (2002) *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge University Press, Cambridge.
- [29] 金艳, 孙杭. 外语课堂评估研究(2007-2018): 回顾与展望[J]. 东北师大学报(哲学社会科学版), 2020(5): 166-173.
- [30] 蔡基刚. 高等教育国际化背景下的外语教学评价体系调整[J]. 外语电化教学, 2013(1): 3-8.