

· 统计、测量与方法 ·

基于认知诊断测评的个性化补救教学效果分析： 以“一元一次方程”为例^{*}

王立君 唐 芳 詹沛达^{**}

(浙江师范大学教师教育学院心理学系, 金华, 321004)

摘 要 基于“为学习而测评”理念,以促进学生学习为目的,本研究进行了基于认知诊断测评的个性化补救教学效果分析。首先,以“一元一次方程”章节为例,编制两份平行的认知诊断测评试卷。然后,通过对不同地区(城市和农村)七年级学生的施测与分析,发现城市学生对属性的掌握情况优于农村学生对属性的掌握情况。之后,选择农村学生为补救对象,通过对比基于认知诊断测评和传统教学两种个性化补救教学的效果,发现两种补救教学方法均能提高学习成绩,但前者的补救效果显著优于后者的。总之,本研究结果表明采用基于认知诊断测评的个性化补救教学能够有效促进学生学习,为实践者应用认知诊断测评促进学生学习提供了实践依据。

关键词 认知诊断 个性化补救教学 为学习而测评 认知诊断模型

1 引言

“学习评价的主要目的是全面了解学生学习的过程和结果。评价既要关注学生学习结果,也要重视学习的过程。教师教学应该以学生的认知发展水平和已有的经验为基础,面向全体学生,注重启发式的教学方法,做到因材施教”(中华人民共和国教育部,2012)。“因材施教”的基本前提是要对学生的现有的学习状况(知识掌握情况)及其在不同时间点上的发展情况(知识掌握程度是否增加?)有一个相对客观和准确的了解(詹沛达,潘艳方,李菲茗,待发表)。“有意义接受学习”理论(Ausubel, Novak, & Hanesian, 1968)认为影响学生学习的最重要因素是学生已有的认知结构,认为有意义接受学习是学生将新知识纳入已有认知结构的过程。因此,客观量化学生学习现状对促进学习具有重要作用。同时,基于“为学习而测评(AFL)”理念(Wiliam,

2011),以促进学生学习为目的,客观量化学生学习现状并提供诊断反馈的测评模式日益受到人们的重视(詹沛达等,待发表)。AFL理念强调反馈对促进学习的重要性,认为在教学过程中应及时测评学生的学习情况并给予相应的反馈,进而促进学生学习。

随着心理计量学的发展,认知诊断测评(CDA)在近些年得到国内外学者的广泛关注。CDA旨在为每位学生提供个性化的诊断反馈报告,进而明确学生的学习现状,并为后续学生的自我补救和教师的有针对性补救教学提供参考依据(李令青,韩笑,辛涛,刘彦楼,2019; Zhan, Jiao, Liao, & Li, 2019)。补救教学是在对学生学习现状诊断的基础上实施的一系列诊疗式教学活动,其通过“测评—施救—再测评”的循环模式缩小不同学生之间的学习水平差异(武小鹏,张怡,2018)。基于CDA的补救教学的成效对实现有效促进学生学习这一最终目标至关重要。

^{*} 本研究得到全国教育科学规划教育部重点课题(主观题的多分属性认知诊断模型开发及其在物理测验中的应用)(课题批准号:DBA150236)的资助。

^{**} 通讯作者:詹沛达。E-mail: pdzhan@gmail.com

DOI:10.16719/j.cnki.1671-6981.20200630

因此，在 CDA 中针对个性化补救教学的效果分析是一个非常值得关注的议题。对此，本文拟尝试进行基于 CDA 的个性化补救教学效果分析，以期证实基于 CDA 的个性化补救教学在促进学生学习方面的显著成效，并为实践者应用 CDA 来促进学生学习提供依据。

全文撰写结构如下：首先，本研究将以七年级数学“一元一次方程”章节为例，编制两份平行的《“一元一次方程”认知诊断测评试卷》，并通过预测验对试卷质量进行把控。其次，把所编制的试卷分别在城市和农村学校进行施测，进而对比城市和农村学生对所考查属性掌握情况的差异。然后，对属性掌握情况较差的学生分组进行基于 CDA 的和基于传统教学的个性化补救，进而实现个性化补救教学的效果分析。最后是总结和讨论。

2 认知诊断测验的编制

2.1 属性和属性层级关系的确认

首先，基于我国《义务教育数学课程标准（2011 版）》的要求，对考试大纲、教材及配套的教师用书、各章节习题和各省测试题进行汇总。然后，采用回顾文献法，参考 Tatsuoaka, Corter 和 Tatsuoaka（2004）的认知属性框架，认为数学认知诊断研究中主要包含三种属性（内容属性、过程属性、技能属性）：内容属性包括各种基本概念，过程属性包括代数规则的运用和分析问题，技能属性包括解决问题。本研究内容属性主要体现在属性 A1（方程的基本概念）中；过程属性主要体现在 A2（代数规则）和 A3（解方程）中；技能属性主要体现在 A5（方程的简单应用）和 A6（方程的复杂运用）中。在康春花、辛涛和田伟（2013）的“小学数学应用题”认知属性框架中，提出实际运用需要识别隐含条件（A4）这一属性，且将比较复杂的属性（行程问题）进行细分（分为一般行程问题和复杂行程问题），从而更好的对学生进行诊断。本研究参考此文将方程的运用也分为属性 A5 和 A6。最后，结合专家判断法，经数学

学科专家、一线数学教师和心理计量学专家的共同讨论，最终确定了本研究的认知属性框架（见表 1）和属性层级结构（见图 1）。在专家确定的过程中，如果出现意见不统一的情况，需进行再一次讨论，最后直到意见统一或大多数专家（超过三分之二）都同意为止。

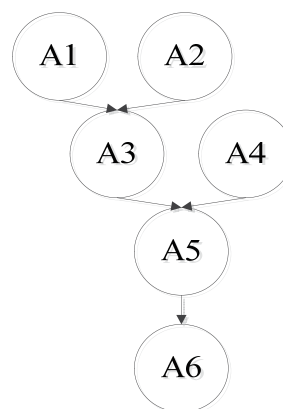


图 1 属性层级结构图

2.2 题目编制

首先，基于属性层级结构得出反映属性间关系的可达（R）矩阵。然后根据扩张算法得到理想掌握模式和题目考查模式，其中题目考查模式为理想掌握模式删除全 0 模式。为保证所编制的测验能够对所有的理想掌握模式进行分类，本研究中的测验 Q 矩阵中包含一个 R 阵（丁树良，杨淑群，汪文义，2010）。然后，每个属性至少被考查两次，以保证诊断的正确率。测验初期共包含 37 道题目，经过预测验的初步筛选并考虑到正式施测时测验时长对学生造成的影响，本研究最终将题目总数控制在 24 题，并划分为两套平行试卷（每套试卷长度为 12 题），测验均采用 0-1 计分方式，即测验满分为 12 分。其中 A 卷用于正式测验中的前测，B 卷用于干预后的后测。测试卷 A 和 B 的 Q 矩阵如表 2 所示，两个 Q 矩阵本质上是一致的，只是在题目顺序（i.e., 行的顺序）上有所差别。

表 1 七年级“一元一次方程”章节对应认知属性框架

属性编码	范围界定
A1 方程的基本概念	一元一次方程、等式、元、次数
A2 代数规则	四则运算、合并同类项和移项、去括号、去分母
A3 解方程	求出使方程的等号左右两边相等的未知数的值
A4 识别隐含条件	从给出信息中找出隐含的条件
A5 方程的简单应用	列方程解决实际问题。如：比率问题，销售盈亏问题
A6 方程的复杂运用	列方程解决实际问题。如：追及问题，流水行船问题

表2 “一元一次方程”认知诊断测评A卷和B卷的测验Q矩阵

题目	A 卷						B 卷					
	A1	A2	A3	A4	A5	A6	A1	A2	A3	A4	A5	A6
1	1	1	0	0	0	0	1	1	1	0	0	0
2	1	0	0	0	0	0	1	1	0	0	0	0
3	1	1	1	0	0	0	1	1	1	1	0	0
4	1	1	1	1	0	0	1	0	0	0	0	0
5	1	1	1	1	1	0	1	1	1	1	1	0
6	0	1	0	0	0	0	0	1	0	0	0	0
7	0	1	0	0	0	0	0	1	0	0	0	0
8	0	0	0	1	0	0	0	0	0	1	0	0
9	1	1	1	0	0	0	1	1	1	0	0	0
10	1	1	1	0	0	0	1	1	1	0	0	0
11	1	1	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1	1	1

3 预测验

为评估题目质量和两份试卷是否为平行卷，在正式测验之前先进行预研究。

3.1 测与被试

被试为某中学（城市）七年级在读学生。研究者利用所编制的认知诊断测评工具对学生进行团体施测，测验初期共包含37道题目，施测时间为120分钟。预测验共发放78份测试卷，回收有效测试卷为74份，有效率为95%。

3.2 分析与结果

依据经典测量理论（CTT）来判断测验和题目的整体质量。首先根据各题目的难度、区分度等，选取难度范围在[.2, 1.0]和区分度范围在[.2, 1.0]之间的题目24道。然后，针对这24题做后续的分析。本测验的Cronbach' α 系数为.79，信度良好。题目难度范围在[.27, .95]之间，题目区分度范围为[.25, 1.00]。其中，本次测验难度系数在.5左右和区分度大于.4的题目均占总题目一半以上。由于认知诊断测验意在诊断出学生的属性掌握情况，所以困难的知识点也必须有考察的题目，这样就可能导致有的题目区分度较小。实际上认知诊断测验不能一味追求高区分度。另外，为探究两份试卷是否满足平行

试卷的要求，表3呈现了学生在A卷和B卷上数学成绩的平均数（ M ）、标准差（ SD ）、题目难度（ P ）和题目区分度（ D ），发现两份试卷成绩的各描述性统计指标基本一致。另外，对两份试卷的数学成绩进行配对样本 t 检验（ $t = -.20, p > .05$ ），发现总体均值差异不显著。最后，鉴于两份试卷测量内容相同，包含相同数量的题目、题型和本质上相同的测验Q矩阵（题目呈现顺序存在差异），且两份试卷的复本信度（ $r = .86, p < .01$ ），即两份试卷相关显著。可认为本研究所编制的两份问卷满足平行试卷的要求。

4 基于认知诊断测评的个性化补救教学效果分析

4.1 正式测验阶段

4.1.1. 被试

被试为两所中学（城市85人，与预测验同一学校，不同班级；农村244人）七年级在读学生。研究者利用所编制的认知诊断测评工具对学生进行团体施测，测试内容为A卷，测试时间为40分钟，共发放问卷335份，回收有效问卷329份，有效率为98%。

表3 AB卷描述性统计结果

	学习成绩		
	$M \pm SD$	P	D
A 卷	7.81 \pm 2.29	.64	.46
B 卷	7.84 \pm 2.10	.64	.42

注： M = “平均数”； SD = “标准差”； P = “难度”； D = “区分度”。

4.1.2 分析与结果

选用合适的 CDM 是获得合理数据分析结果（或诊断反馈结果）的必要条件。对此，本研究首先选用 3 个常用的 CDM 来分析数据，包括 DINA 模型（Junker & Sijtsma, 2001）、DINO 模型（Templin & Henson, 2006）和 GDINA 模型（de la Torre, 2011）。然后，基于模型 - 数据拟合指标，选出与数据相对拟合度最优的模型作为后续数据分析的模型。

本研究采用 R 软件中的 CDM 包（Robitzsch, Kiefer, George, & Uenlue, 2019）进行数据分析。表 4 给出了三个模型的模型 - 数据拟合指标，AIC 与 BIC 为相对拟合指标，MADcor 与 SRMSR 为绝对拟合指标。综合以上四个指标，发现 DINA 模型的相对拟合情况较好，表明本测验中属性之间可能满足连接缩合规则（Maris, 1995）。因此，下文将详细呈现 DINA 模型的分析结果。

表 5 呈现了基于 DINA 模型的学生诊断分类结果，其中属性掌握模式 (000000) 和 (111111) 的人数最多，分别占 30.83% 和 36.95%。其次是属性掌握模式 (010000)，占 11.08%。其他属性掌握模式占比

均比较少。另外，表 6 呈现了全体学生对 6 个属性的掌握比例，从高到低依次为 A1、A2、A4、A3、A5 或 A6。

表 7 呈现了城市学生 (85 人) 和农村学生 (244 人) 对各属性的掌握比例。可以看出，城市学生对 6 个属性的掌握比例均高于农村学生。此外，农村和城市两个地区学生数学成绩的均值（标准差）分别为 3.38 (2.29) 和 7.94 (2.30)，进一步对两地区学生的数学成绩进行独立样本 t 检验，发现两者总体均值差异显著 ($t = -15.80, p < .001$)。该结果表明，城市学校学生的数学成绩显著高于农村学校学生的数学成绩。

4.2 个性化补救阶段

4.2.1 补救过程

正式测验结果表明，相对于城市学生而言，农村学生对各属性的掌握相对较差。为便于补救教学的实施，本研究将补救教学对象定为农村学生。根据诊断结果和学生平时成绩进行有目的的随机抽样 22 名农村学生，平均分为两组：基于 CDA 的个性化补救组（简称为“认知诊断组”）和传统补救教

表 4 模型 - 数据拟合指标

模型	AIC	BIC	MADcor	SRMSR
DINA	4304.61	4437.47	.058	.078
DINO	4341.21	4474.07	.063	.081
GDINA	4468.67	5315.19	.037	.052

表 5 基于 DINA 模型的诊断分类结果

属性掌握模式	概率	人数
000000	30.83%	101
100000	6.15%	20
010000	11.08%	36
111000	6.56%	22
000100	1.78%	6
100100	5.26%	17
111100	1.38%	5
111111	36.95%	122

表 6 全体学生对六个属性的掌握比例

A1	A2	A3	A4	A5	A6
56.30%	55.98%	44.89%	45.37%	36.95%	36.95%

表 7 城市学生和农村学生对六个属性的掌握比例

	A1	A2	A3	A4	A5	A6
农村	70.13%	49.74%	40.47%	35.95%	25.96%	25.96%
城市	91.14%	71.55%	70.41%	76.89%	59.53%	59.53%

学组（简称为“传统教学组”）。为避免不同干预形式（一对一干预和团体干预）所带来的差异，两组学生都将接受一对一的个性化补救教学；不同在于，认知诊断组中，将对学生未掌握的属性进行逐个补救，而传统教学组中，只对学生讲解错题本身，而非特定属性。对认知诊断组中的学生，每个属性补救方法分为以下几个步骤（认知诊断个性化补救材料可向作者索取）：第一步，向学生解释该属性相关的概念；第二步，给出考察该属性的例题，向学生进行讲解；第三步，给出考察该属性的问题，学生自行解答；第四步，向学生讲解第三步的问题。两组学生补救总时长均为 10h，每个学生均在补救两天后，使用 B 卷对其进行后测。

4.2.2 结果分析

首先，采用 CTT 分析方法，认知诊断组和传统教学组学生补救前后测数学成绩的描述统计如表 8 所示。认知诊断组和传统教学组学生的学习成绩前测无显著性差异 ($t = -.079, p > .05$)。以补救方法（认知诊断/传统教学）、测试时间点（前测/后测）为自变量，以数学成绩为因变量进行两因素重复测量方差分析。结果见表 9，补救方法的主效应显著 ($p < .05$)，表明补救方法对学生学习成绩有显著影响；测试时间点的主效应显著 ($p < .001$)，表明在两种补救方法下，学生前后测学习测成绩均有显著性差异；补救方法和测试时间的交互效应显著 (p

$< .001$)，进一步进行简单效应分析发现，认知诊断补救效果显著好于传统教学补救。

然后，采用 DINA 模型对学生的属性掌握情况进行分析。由于个性化补救阶段样本量比较小，为了减小 DINA 模型分析的误差，本研究将补救后学生的数据与之前预测阶段 B 卷的数据一起分析（具体数据分析流程见图 2，另：参数估计 1 也可以加上预测阶段 A 卷的数据）。从 Basokcu (2014) 研究中可知，当样本量 $N=100$ 时，与样本量为 $N=1000$ 时其中相同的 100 人的分类一致率为 91.4%，本研究中两次参数估计的样本量均接近或大于 100，所以使用 DINA 模型估计是合理的。

分析结束后将补救前和补救后的学生看作不同个体进行分析，结果如表 10 所示。传统教学组学生在 A1、A2 和 A3 四个属性上掌握比例均有增加，在 A5、A6 两个属性上掌握比例没有变化；而认知诊断组学生在 6 个属性上的掌握比例均有增加。

5 总结、讨论与展望

5.1 总结

基于 AFL 理念，以促进学生学习为目的，本研究进行了基于 CDA 的个性化补救教学效果分析。首先，本研究以“一元一次方程”章节为例，编制两份平行的认知诊断测评试卷。预测阶段结果表明两份试卷质量较好且满足平行试卷的要求。然后，将

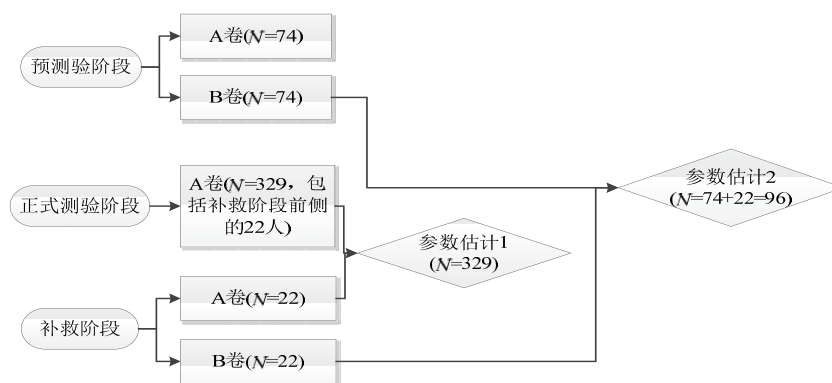


图 2 补救阶段数据分析图

表 8 两种补救方法下学习成绩的描述统计 ($M \pm SD$)

补救方法	学习成绩	
	前测	后测
认知诊断	3.45±2.84	9.27±1.95
传统	3.55±2.54	4.36±2.58

注： M = “平均数”； SD = “标准差”

表 9 学习成绩在补救方法和测试时间点上的差异检验

	学习成绩	
	$F(1)$	p
补救方法	5.71	< .05
测试时间点	90.94	< .001
补救方法×测试时间点	51.62	< .001

表 10 补救前后测各属性掌握比例

	传统		个性化	
	前测	后测	前测	后测
A1	36.36%	72.73%	27.27%	100.00%
A2	36.36%	63.64%	36.36%	100.00%
A3	27.27%	45.45%	18.18%	100.00%
A4	9.09%	9.09%	9.09%	90.91%
A5	0.00%	0.00%	0.00%	72.73%
A6	0.00%	0.00%	0.00%	63.64%

所编制的试卷分别在城市和农村学校进行施测，发现城市学生对属性的掌握情况优于农村学生对属性的掌握情况。之后，通过对比基于 CDA 的和基于传统教学的个性化补救方法发现，两种补救方法都能提高学生的成绩且基于 CDA 的个性化补救教学效果显著优于基于传统的补救教学。总之，本研究结果表明采用基于 CDA 的个性化补救教学能够有效促进学生学习。

5.2 讨论

5.2.1 总体学生的属性掌握情况

从学生的属性掌握模式归类情况来看（见表 5），模式（000000）和（111111）所占的人数最多，分别占比 30.83% 和 36.95%。全都没掌握和全都掌握占比较大这一情况从侧面表明了各知识点联系紧密。此外，属性掌握模式（010000）人数较多，占 11.08%。这在一定程度上可以表明属性 A2 比较简单。从属性层级结构看，A2 属于比较基础的属性，学生掌握得较好是符合属性层级结构的。从表 6 可以看出，学生认知属性掌握比例从高到低依次为 A1、A2、A4、A3、A5 或 A6，进一步验证了属性层级关系的合理性：属性 A1 和 A2 是基础属性，属性 A3 和 A4 是中层属性，而 A5 和 A6 为高层属性。该研究结果也与实际教学中知识点难度情况相符。

5.2.2 不同地区的学生属性掌握情况

由不同地区（城市、农村）学生的属性掌握情况（见表 7）可知，城市学校的学生在 6 个属性上的掌握情况均比农村学校的学生要好。这可以从城市学校与农村学校的教学环境、教师资源和生源质

量等多方面来解释，所以加大农村教育资源投入是非常必要的。基于 CDA 的个性化补救教学能够让学生了解自己的学习情况。特别是对于农村学校基础比较薄弱的学生，他们大多并不知道自己对知识的掌握情况，所以了解到自己所缺失的部分能够让教师或学生自己进行针对性的补救，从而辅助教师的教

5.2.3 认知诊断补救情况

结合表 8、表 9 可以看出，认知诊断组学生后测成绩显著高于传统教学组，且认知诊断组补救后较传统教学组有更显著的提升。从表 10 中可以看出，认知诊断组的学生补救后各属性掌握比例较传统教学组增加幅度更大。具体而言，认知诊断组的学生在补救后 6 个属性掌握比例均有增加，传统教学组在 A1、A2 和 A3 三个属性上掌握比例均有增加，但在 A4、A5、A6 三个属性上掌握比例没有任何变化。因为属性 A1、A2 是比较基础的属性，学生掌握起来比较简单，属性 A3 虽然是中层属性，但从属性层级结构可知，A3 可以通过提升 A1 和 A2 的掌握程度来掌握。所以在经过笼统的传统教学补救的学生在 A1、A2 和 A3 三个属性上掌握比例均有增加。A4、A5、A6 三个属性为中层或高层属性，学生掌握起来比较困难，只是针对错题进行讲解不足以让学生融会贯通，需要对其进行概括性的讲解。

在传统教学中，教师在进行章节小测后一般是对学生出错较多的题目进行重点讲解，而出错较少的题目可能会一笔带过。但每个学生对知识点的掌握情况是不一样的，对于简单的知识点，也有一部分学生没有掌握，教师的一笔带过可能会让部分学

生错失学习机会,之后学习困难的知识点可能会更加吃力。教师对难题做重点讲解,但如果没有简单知识点的铺垫和对其归类总结(归类成属性,概括性讲解)的话,学生可能会听不懂或者只学会了一道题,而不会举一反三。结合实际情况,成绩较好的学生大多数可以总结出自己哪块知识点比较欠缺,会进行课后自行补救学习。但对于成绩比较一般或者较差的学生,该类学生一般只知道自己很多都不会,但具体不会在哪里又说不上来。综上,采用基于CDA的个性化补救教学,让学生了解自己对知识的掌握情况是十分有必要的。

5.3 研究不足与展望

本研究为应用型研究,旨在证实基于CDA的个性化补救教学在促进学生学习方面具有显著成效。尽管研究结果已经可以为实践者应用CDA来促进学生学习提供实践依据,但本研究仍存有不足,有待在后续研究中尝试解决。第一,由于个性化补救阶段样本量较小,可能导致后测分析不够准确,今后可尝试使用基于CDA的团体补救教学来增加样本量;第二,如正文所述,选用合适的CDM是获得合理诊断反馈结果的必要条件。尽管DINA模型能够相对较好地拟合本研究中的数据,但并不表示该模型适用于所有题目。在不同的题目中属性之间的关系可能并不相同(高旭亮,汪大勋,蔡艳,涂冬波,2018; Ma, Iaconangelo, & de la Torre, 2016; Zhan, 2020)。因此,在后续研究中可尝试使用混合模型来分析数据。第三,为探究补救教学的成效,纵向研究是必需的。本研究仅采用了包含两个测验时间点的前后测设计,尚无法刻画学生的学习轨迹。因此,在后续研究中可尝试使用包含更多测验时间点的纵向测评设计(Zhan et al., 2019)。第四,本研究在前后测中使用了平行试卷,但现实中并没有完全平行的两份试卷。因此,后续可尝试采用锚题设计的试卷来链接各测评时间点。第五,本研究仅采用了纸笔测验和人工补救教学,在“互联网+”测评(张华华,汪文义,2016)时代里,需要更智能化的诊断和补救工具。因此,在后续研究中可尝试使用基于计算机(网络)的个性化补救教学和自适应学习来促进学生学习。

参考文献

戴海琦.(2010).心理测量学.北京:高等教育出版社.

- 丁树良,杨淑群,汪文义.(2010).可达矩阵在认知诊断测验编制中的重要作用.《江西师范大学学报(自然科学版)》,34(5),490-494.
- 高旭亮,汪大勋,蔡艳,涂冬波.(2018).认知诊断模型的比较及其应用研究:饱和模型、简化模型还是混合方法.《心理科学》,41(3),727-734.
- 康春花,辛涛,田伟.(2013).小学数学应用题认知诊断测验编制及效度验证.《考试研究》,6,24-43.
- 李令青,韩笑,辛涛,刘彦楼.(2019).认知诊断评价在个性化学习中的功能与价值.《中国考试》,1,40-44.
- 武小鹏,张怡.(2018).基于认知诊断视角的数学补救教学.《教育测量与评价》,7,33-40.
- 詹沛达,潘艳方,李菲茗.(待发表).面向“为学习而测评”的纵向认知诊断模型.《心理科学》.
- 张华华,汪文义.(2016).“互联网+”测评:自适应学习之路.《江西师范大学学报(自然科学版)》,40(5),441-455.
- 中华人民共和国教育部.(2012).义务教育数学课程标准(2011年版).北京:北京师范大学出版社.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1968). *Educational psychology: A cognitive view (Vol. 6)*. New York: Holt, Rinehart and Winston.
- Basokcu, T. O. (2014). Classification accuracy effects of Q-matrix validation and sample size in DINA and G-DINA models. *Journal of Education and Practice*, 5(6), 220-230.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76(2), 179-199.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3), 258-272.
- Ma, W. C., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, 40(3), 200-217.
- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60(4), 523-547.
- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2019). CDM: Cognitive Diagnosis Modeling. *R Package Version 7.3-17*.
- Tatsuoka, K. K., Corter, J. E., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, 41(4), 901-926.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3), 287-305.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Wu, H. M. (2019). Online individualised tutor for improving mathematics learning: A cognitive diagnostic model approach. *Educational Psychology*, 39(10), 1218-1232.
- Zhan, P. D. (2020, February 12). *Cognitive diagnostic mixture modeling for adaptively identifying coexisting condensation rules*. <https://doi.org/10.31234/osf.io/qwx2m>
- Zhan, P. D., Jiao, H., Liao, D. D., & Li, F. M. (2019). A longitudinal higher-order diagnostic classification model. *Journal of Educational and Behavioral Statistics*, 44(3), 251-281.

Effect Analysis of Individualized Remedial Teaching based on Cognitive Diagnostic Assessment: Taking “Linear Equation with One Unknown” as An Example

Wang Lijun, Tang Fang, Zhan Peida

(Department of Psychology, College of Teacher Education, Zhejiang Normal University, Jinhua, 321004)

Abstract Based on the idea of "assessment for learning" and aiming at promoting students' learning, the assessment pattern of objectively quantifying the learning status and providing diagnostic feedback has been increasingly valued. Cognitive diagnostic assessments (CDAs) aim to provide individualized diagnostic feedback for each student, clarify students' learning status, and provide a reference for follow-up students' self-remedy and teachers' targeted remedial teaching. To achieve the ultimate goal of effectively promoting students' learning, the effectiveness of remedial teaching based on cognitive diagnosis is very important.

However, currently, limited studies focus on evaluating the effectiveness of remedial teaching. One of the main reasons is that the cross-sectional design is adopted by most CDAs, which means that only one assessment is done at a specific testing time point. This issue may also reflect on the development of existing cognitive diagnosis models (CDMs). Although various CDMs have been proposed, most of them are only applicable to cross-sectional data analysis.

In recent years, the longitudinal data collected from the CDAs throughout the learning process provides researchers with the chance to develop learning diagnosis models that can be adapted to track individual growth over time and to evaluate remedial teachings' effectiveness. Compared to the cross-sectional CDA, longitudinal CDA is more helpful when aiming to promote students' development. Under such a background, in longitudinal CDAs, the effect analysis of individualized remedial teaching is a very interesting topic.

To explore the effectiveness of individualized remedial teaching based on the longitudinal CDA, this research is divided into four stages: test development stage, preliminary testing stage, formal testing stage, and individualized remedial teaching stage. In test development stage, two parallel cognitive diagnostic tests are compiled by taking the chapter of "linear equation with one unknown" as an example, including the Volume A and Volume B. Two parallel tests both require six attributes and have the same Q-matrix (the order of the rows in two Q-matrices is different), including (A1) basic concept of linear equation; (A2) algebra rules; (A3) solving linear equation; (A4) recognition of implicit conditions; (A5) simple application of the linear equation; and (A6) complex application of the linear equation. In the preliminary testing stage, there are 74 seventh-grade students taking part in the testing. The two tests are combined into one and the test time is 80 minutes. Based on the analysis via classical test theory, the results indicate that two tests both have good quality and meet the requirements of parallel tests. In the formal testing stage, there are 329 seventh-grade students (85 in the urban area and 244 in the rural area) taking part in the testing. Volume A is used and the test time is 40 minutes. Three CDMs, including the DINA, DINO, and GDINA models, are used to fit the data. The model-data fit indices (AIC and BIC) preferred the DINA model as the best fitting model. Further, the results of the DINA model found that the mastery of attributes by urban students is better than that by rural students. In the individualized remedial testing stage, 22 rural students are sampled and divided into two groups on average: the CDA-based remedial teaching group and the traditional remedial teaching group. In the first group, students receive one-to-one remedies from the teacher according to the diagnostic feedback that focuses on non-mastery attributes. By contrast, students in the second group receive one-to-one remedies from the teacher according to the incorrect item responses. By comparing the effect of CDA-based individualized remedial teaching and traditional remedial teaching, it is found that both remedial methods can improve students' performance and the effect of CDA-based individualized remedial teaching is significantly better than that of traditional remedial teaching ($F(1) = 51.62, p < .001$).

Overall, the results of this study indicate that the CDA-based individualized remedial teaching can effectively promote students' learning. The results of this study also provide a practical basis for practitioners to use CDA to promote students' learning.

Key words cognitive diagnosis, individualized remedial teaching, assessment for learning, cognitive diagnostic model