

常见的认知诊断模型及其比较

范士青 刘华山

【摘要】认知诊断模型是连接学生的内在属性掌握模式和外在题目反应之间的桥梁,是教育评估领域一个新的发展方向。经过40多年的发展,认知诊断模型目前有60种左右,DINA、DINO、NIDA、NIDO、R-RUM和C-RUM是最常见的模型。这六种认知诊断模型适用于题目计分、属性掌握表征和题目对属性的考查都采用二值方式的情况,都以概率的形式表达了特定的知识状态在特定题目上正确作答的概率。这六种模型的区别主要表现在三个方面:①依据属性之间的关系,DINA、NIDA和R-RUM属于非补偿模型,DINO、NIDO和C-RUM属于补偿模型;②DINA和DINO的失误参数和猜测参数定义在题目水平上,NIDA和NIDO的参数定义在属性水平上;③六个模型中除NIDO和C-RUM采用的是logit函数形式,其他都采用了恒等的形式。梳理出常见的认知诊断模型及其背后蕴含的思想方法、模型结构及各自特点,能够为认知诊断评估的进一步推广提供助力。

【关键词】认知诊断模型;补偿模型;非补偿模型;参数设置

【中图分类号】G449.1 【文献标识码】A 【DOI编码】10.16518/j.cnki.emae.2015.07.002

认知诊断评估是教育评估领域一个新的发展方向,它主要是通过对学生在的一组精心设计的题目上的作答反应进行分析,推测学生在测验所涉及的一组知识或技能(属性)上的掌握情况,从而为学生个别辅导、教师教学反馈提供依据。认知诊断评估的关键是构建反映学生问题解决过程的Q矩阵并选用适当的认知诊断模型。认知诊断模型是一种通过描述学生的属性掌握情况与其题目反应之间关系以实现认知诊断功能的测量模型,属性是测验所考察的知识或技能的统称。经过40多年的发展,认知诊断模型目前有60种左右。^[1]本研究介绍并比较了DINA、DINO、NIDA、NIDO、R-RUM和C-RUM这六种当前最常见的具体认知诊断模型,及其背后蕴含的思想方法、模型结构及各自特点,并以实例和表格梳理出不同认知诊断模型之间的基本区别。

一、认知诊断模型的假设

认知诊断模型以Q矩阵理论为基础,模型之

间的差异反映了不同的认知假设。认知诊断模型依据属性的个数可以分为单维模型和多维模型,目前以多维模型为主;依据潜变量之间的相互作用关系可以分为补偿模型和非补偿模型;依据模型对学生属性掌握表征方式的假定,分为连续型和离散型两种,采用连续型属性表征方式的模型被称为潜在特质模型,采用顺序或二分的离散性表征方式的模型被称为潜在分类模型^[2];依据模型是否假定测验考查的所有属性均被包含在Q矩阵中,可以对Q矩阵的完备性作出划分,绝大部分模型假定Q矩阵是完备的,假定其不完备的模型常用一个连续变量来刻画Q矩阵以外的残余技能;依据模型是否考虑不同学生同一题目上可能采用了不同的策略,可以把模型划分为单一策略模型和多重策略模型。常见的认知诊断模型认知假设如表1所示。

二、认知诊断模型的公式和符号体系

认知诊断模型的基本设想是学生对题目所

本研究为湖北省教育厅人文社科青年项目“小学生减法计算错误的认知诊断与模型比较研究”(项目编号:14Q120)阶段性成果。

范士青/华中师范大学心理学院博士生,湖北第二师范学院教育科学学院教师,研究方向为教育心理学与学校心理辅导。(武汉 430205)

刘华山/华中师范大学心理学院教授,博士生导师。通讯作者。

常见的认知诊断模型及其比较

表 1 主要的认知诊断模型认知假设分类一览表

模型	维度 个数	属性间 补偿性	属性掌 握表征	Q 矩阵 完备性	多重 策略
LLTM	单维	补偿	连续	是	否
MIRT-C	多维	补偿	连续	是	否
MIRT-NC	多维	非补偿	连续	是	否
GLTM	多维	非补偿	连续	否	否
MLTM	多维	非补偿	连续	否	是
DINA	多维	非补偿	离散	是	否
NIDA	多维	非补偿	离散	是	否
DINO	多维	补偿	离散	是	否
NIDO	多维	补偿	离散	是	否
HYBIRD	多维	非补偿	均可	否	是
R-RUM	多维	非补偿	离散	否	否
C-RUM	多维	补偿	离散	否	否
RSM	多维	非补偿	离散	是	否

考查的属性掌握的个数越多,则其答对该题的可能性就越大。为了梳理出具体认知模型的内在特点,我们选取表 1 中 DINA、DINO、NIDA、NIDO、R-RUM 和 C-RUM 模型加以分析,这些模型在认知诊断评估中得到了较为广泛的应用,其参数设置各有特点,且属性间关系涵盖了补偿和非补偿两种类型。考虑到不同认知诊断模型所采用的符号体系并不一致和一线教师的数理知识基础,我们按照如下方式约定统一的符号体系。

认知诊断评估领域所考查的属性个数为 K ,其中第 k 个属性用 A_k 表示,这些属性在测验编制之前由研究者采用文献综述或认知心理学的方法加以确定,属性的含义及属性之间的层次关系亦应阐明。根据这些属性编制一个题目个数为 J 的认知诊断测验,各个题目对属性的考查情况不尽相同。为了描述这种考查情况,研究者多采用一个 $J \times K$ 的 Q 矩阵来标定题目对属性的考查情况,矩阵行向量 q_j 反映了题目 j 对 K 个属性的考查情况,向量元素 q_{jk} 为 1 时表示题目 j 考查了属性 k ,0 表示没有考查属性 k ,题目 j 所考查的属性个数用 K_j^* 来表示。设有 I 个学生参加了该测验,学生 i 对 K 个属性的掌握模式用向量 α_i 表示,其元素 α_{ik} 多用 1、0 表示学生 i 对属性 k 掌握与否,少量模型采用等级或连续变量反映不同的

属性掌握水平。学生 i 在题目 j 上反应的计分变量 X_{ij} 用 1、0 表示对和错;若为多值计分,则 X_{ij} 可取任意非负数值。认知诊断评估就是基于已知的 Q 矩阵和学生在不同题目上的作答反应来考查学生对 K 个属性的掌握情况。当属性掌握表征采用离散型的表达方式时,学生对 K 个属性掌握的可能情况总是有限的。例如,在一个属性个数 $K=3$ 的认知模型中,在不考虑属性间层次关系且属性掌握模式采用二值表征的前提下,所有可能的知识状态有 $2^3=8$ 种,如表 2 所示。学生的掌握模式 α_i 虽是未知的,但必定且只能是这 8 种知识状态中的一种。依此认知模型编制出的认知诊断测验中,考虑到任何属性都不考查的题目并不符合认知诊断要求,因此题目对属性的考查情况 q_j 只有 7 种独特的模式,这些模式组合起来即为 Q 矩阵,如表 3 所示。

表 2 学生所有可能的知识状态($K=3$)

KS	A1	A2	A3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

表 3 七个题目对应的 Q 矩阵($K=3$)

题目(T)	A1	A2	A3
1	0	0	1
2	0	1	0
3	0	1	1
4	1	0	0
5	1	0	1
6	1	1	0
7	1	1	1

表 4 列出了这 6 种认知诊断模型的公式、参数的定义及学生的属性掌握模式 α_i 对题目 j 上作答计分变量的影响。为了简洁地说明不同认知诊断模型之间的关系且不失一般性,我们从表 3 中 K_j^* 取值分别为 1、2 和 3 的题目中分别选取 T4、T6 和 T7,依照表 4 所示的 6 种认知诊断模型、表 2 中的 8 种不同知识状态,对应到这 3 个题目上的正确作答概率分别如表 5~表 8 所示。我们依据概率取值将知识状态分为若干个组,各组之间的概率不同,同组内部不同知识状态对应的概率相等。分组个数取决于模型的假设和题目考查的属性个数 K_j^* ,在属性采用二值表征的前提

常见的认知诊断模型及其比较

表 4 六种主要的认知诊断模型及参数定义		
模型	公式	参数及含义
DINA ^[3]	$P(X_{ij}=1 \alpha_i)=(1-s_j)g_j^{\xi_{ij}(1-\xi_{ij})}$	$s_j=P(X_{ij}=0 \xi_{ij}=1)$
	$\xi_{ij}=\prod_{k=1}^k\alpha_{ik}^{q_{jk}}$	$g_j=P(X_{ij}=1 \xi_{ij}=0)$
NIDA ^[4]	$P(X_{ij}=1 \alpha_i)=\prod_{k=1}^k[(1-s_k)^{\alpha_{ik}(1-\alpha_{ik})}g_k^{q_{jk}}]$	$s_k=P(Y_{ijk}=0 \alpha_{ik}=1, q_{jk}=1)$
		$g_k=P(Y_{ijk}=1 \alpha_{ik}=0, q_{jk}=1)$
R-RUM ^[5]	$P(X_{ij}=1 \alpha_i)=\pi_j^*\prod_{k=1}^k r_{jk}^{*q_{jk}(1-\alpha_{ik})}$	$\pi_j^*=\prod_{k=1}^k P(Y_{ijk}=1 \alpha_{ik}=1)^{q_{jk}}$
		$r_{jk}^*=\frac{r_{jk}}{\pi_{jk}^*}=\frac{P(Y_{ijk}=1 \alpha_{ik}=0)}{P(Y_{ijk}=1 \alpha_{ik}=1)}$
DINO ^[6]	$P(X_{ij}=1 \alpha_i)=(1-s_j)\omega_j^{(1-\omega_j)g_j}$	$s_j=P(X_{ij}=0 \omega_j=1)$
	$\omega_j=1-\prod_{k=1}^k(1-\alpha_{ik})^{q_{jk}}$	$g_j=P(X_{ij}=1 \omega_j=0)$
NIDO ^[7]	$P(X_{ij}=1 \alpha_i)=\frac{\exp(\sum_{k=1}^k(\lambda_{ik}\alpha_{ik}+\lambda_{0k})q_{jk})}{1+\exp(\sum_{k=1}^k(\lambda_{ik}\alpha_{ik}+\lambda_{0k})q_{jk})}$	λ_{0k} 学生未掌握属性 k 的情况下正确作答概率的对数发生比
		λ_{ik} 学生掌握了属性 k 时正确作答概率对数发生比的增加值
C-RUM ^[8]	$P(X_{ij}=1 \alpha_i)=\frac{\exp(\sum_{k=1}^k r_{jk}^*\alpha_{ik}q_{jk}-\pi_j^*)}{1+\exp(\sum_{k=1}^k r_{jk}^*\alpha_{ik}q_{jk}-\pi_j^*)}$	$-\pi_j^*$ 学生任何属性均未掌握情况下正确作答概率的对数发生比
		γ_{jk}^* 学生掌握属性 k 时正确作答概率对数发生比的增加值

下,分组数不超过 2^{K_j} 个。

三、认知诊断模型的正确作答概率

DINA 模型是一个参数定义在题目水平上的非补偿模型,通过指标 ξ_{ij} 的取值 1 和 0 表示不同知识状态的学生是否掌握了题目 j 所考查的所有属性,相应的正确作答的概率分别为 $(1-s_j)$ 和 g_j 两种取值。 s_j 和 g_j 为题目水平上的失误和猜测参数,且有 $(1-s_j)>g_j$,其考虑是依照测验实际情况,所有属性均已掌握的学生即使有失误答对的概率也应该高于属性没有掌握的学生猜对题目的概率。由表 5 可见,无论题目所考查的属性个数是多少,其正确作答概率总是被分成两种情况,其题目参数个数总为两个。第一种情况对应的知识状态包含了题目 j 所考查的所有属性;第二种情况对应的每个知识状态中都缺少了题目 j 所考查的属性,缺少的个数不影响作答概率。

DINO 模型是一个参数定义在题目水平上的

补偿模型,适用于学生只要掌握了题目所考查的任意一个属性就有较高作答概率的测验情境,在此基础上多掌握了考查的属性对作答概率不构成影响。模型通过指标 ω_{ij} 取值 1 和 0 表示学生 i 对题目 j 所考查属性是否至少掌握了一个。参数

表 5 DINA 和 DINO 模型下不同知识状态的学生答对不同题目的概率

		DINA 模型			DINO 模型		
KS		T4	T6	T7	T4	T6	T7
		100	110	111	100	110	111
1	000	g_4	g_6	g_7	g_4	g_6	g_7
2	001	g_4	g_6	g_7	g_4	g_6	$1-s_7$
3	010	g_4	g_6	g_7	g_4	$1-s_6$	$1-s_7$
4	011	g_4	g_6	g_7	g_4	$1-s_6$	$1-s_7$
5	100	$1-s_4$	g_6	g_7	$1-s_4$	$1-s_6$	$1-s_7$
6	101	$1-s_4$	g_6	g_7	$1-s_4$	$1-s_6$	$1-s_7$
7	110	$1-s_4$	$1-s_6$	g_7	$1-s_4$	$1-s_6$	$1-s_7$
8	111	$1-s_4$	$1-s_6$	$1-s_7$	$1-s_4$	$1-s_6$	$1-s_7$

注:表中 g_j 和 s_j 分别为定义在题目水平上的猜测参数和失误参数,其下角标数字为题目编号。

常见的认知诊断模型及其比较

s_j 和 g_j 分别是失误参数和猜测参数,且有 $(1-s_j) > g_j$ 。其个数取决于题目个数 J 而与题目所考查的属性个数 k_j^* 无关。从表 5 我们可以看出, DINO 模型和 DINA 模型参数的符号相同, 题目正确作答概率也被分为两组, 但分组方式却明显不同, 因此同一题目在这两个模型下得到的 s_j 或 g_j 也必定是不同的。

NIDA 模型是一个参数定义在属性水平上的非补偿模型。为了定义属性水平上的失误参数 s_k 和猜测参数 g_k , 模型通过指标 Y_{ijk} 取值 1 或 0 来反映一个属性掌握情况为 α_{ik} 的学生能否成功地在题目 j 上应用属性 k 。 s_k 和 g_k 的取值在不同属性间变化而在不同题目间保持不变, 且有 $(1-s_k) > g_k$ 。由表 6 可见, 考虑到属性缺失个数对于正确作答概率的影响, 学生在题目上正确作答概率被分为 $2^{k_j^*}$ 个不同的取值。不同知识状态对应应在题目 j 上的作答概率分组情况依赖于题目所考察的属性个数 k_j^* , 分组个数随 k_j^* 的增加有呈指数级增长的趋势。模型参数的个数取决于整个测验考查的属性个数 K , 与题目个数无关。

RUM 是基于参数识别的目的对统一模型进

行参数重构而提出的模型。该模型有补偿和非补偿两个版本, 非补偿模型也就是融合模型, 它又分为完整版本和精简版本, 其完整版本可以为学生评估出一个离散的属性掌握向量和一个表示残余能力的连续变量 θ_i ; 其精简版本剔除了残余能力。模型参数既有定义在题目水平上的 π_j^* , 又有定义在题目和属性交叉水平上的 r_{jk}^* 。 π_j^* 取值在 0 到 1 之间, 其值越大题目越容易, 表示题目 j 所考查的属性都被掌握的条件下这些属性都能被正确应用的概率, 是以 Q 矩阵为基础的科目难度, r_{jk}^* 为取值在 0 到 1 之间的参数, 其个数等于题目 j 所考查的属性的个数 k_j^* , 定义为学生没有掌握属性 k 但是靠猜测在题目 j 上正确应用该属性的概率, 和掌握了属性 k 的学生且能在题目 j 上正确应用该属性的概率之比, 反映了属性 k 对题目 j 的重要性, 是属性 k 在题目 j 上的区分度, 越小说明 k 越重要^[9], 它在模型中实际起到一个惩罚参数的作用, 即学生没有掌握属性 k 所导致的在正确作答率上的损失。由表 6 我们可以看到, 不同知识状态对应应在题目 j 上的作答概率分组情况取决于其所考查的属性个数 k_j^* 。

NIDO 模型是参数定义在属性水平上的补偿模型, 其特点是采用 logit 函数形式, 即模型参数是题目正确作答概率对数发生比的构成部分, 并不直接对应于正确作答概率。模型中参数 λ_{0k} 和 λ_{1k} 的个数依赖于测验中的属性个数 K , 而与题目个数 J 无关。当且仅当 $q_{jk}=1$ 时 λ_{0k} 和 λ_{1k} 才对题目 j 的正确作答概率产生影响: 当 $\alpha_{ik}=0$ 时, 只有基础率参数 λ_{0k} 会对题目 j 的正确作答概率产生影响; 当 $\alpha_{ik}=1$ 时, 基础率参数 λ_{0k} 和增益参数 λ_{1k} 会对题目 j 的正确作答概率产生影响。由表 7 我们可以看出, 不同知识状态对应应在题目 j 上的作答概率分组情况取决于其所考查的属性个数 k_j^* 。

RUM 模型的补偿模型的版本为 C-RUM, 同样采用 logit 函数形式。 r_{jk}^*

表 6 NIDA 和 R-RUM 模型下不同知识状态的学生答对不同题目的概率

KS	NIDA 模型			R-RUM 模型		
	T4 100	T6 110	T7 111	T4 100	T6 110	T7 111
1 000	g_1	$g_1 g_2$	$g_1 g_2 g_3$	$\pi_4^* r_{41}^*$	$\pi_6^* r_{61}^* r_{62}^*$	$\pi_7^* r_{71}^* r_{72}^* r_{73}^*$
2 001	g_1	$g_1 g_2$	$g_1 g_2 (1-s_3)$	$\pi_4^* r_{41}^*$	$\pi_6^* r_{61}^* r_{62}^*$	$\pi_7^* r_{71}^* r_{72}^*$
3 010	g_1	$g_1 (1-s_2)$	$g_1 (1-s_2) g_3$	$\pi_4^* r_{41}^*$	$\pi_6^* r_{61}^*$	$\pi_7^* r_{71}^* r_{73}^*$
4 011	g_1	$g_1 (1-s_2)$	$g_1 (1-s_2) (1-s_3)$	$\pi_4^* r_{41}^*$	$\pi_6^* r_{61}^*$	$\pi_7^* r_{71}^*$
5 100 (1-s ₁)	$(1-s_1) g_2$	$(1-s_1) g_2 g_3$	$(1-s_1) g_2 g_3$	π_4^*	$\pi_6^* r_{62}^*$	$\pi_7^* r_{72}^* r_{73}^*$
6 101 (1-s ₁)	$(1-s_1) g_2$	$(1-s_1) g_2 (1-s_3)$	$(1-s_1) g_2 (1-s_3)$	π_4^*	$\pi_6^* r_{62}^*$	$\pi_7^* r_{72}^*$
7 110 (1-s ₁)	$(1-s_1) (1-s_2)$	$(1-s_1) (1-s_2) g_3$	$(1-s_1) (1-s_2) g_3$	π_4^*	π_6^*	$\pi_7^* r_{73}^*$
8 111 (1-s ₁)	$(1-s_1) (1-s_2) (1-s_3)$	$(1-s_1) (1-s_2) (1-s_3)$	$(1-s_1) (1-s_2) (1-s_3)$	π_4^*	π_6^*	π_7^*

注 NIDA 模型中 g_k 和 s_k 分别为定义在题目水平上的猜测参数和失误参数, 其下角标数字为题目编号。R-RUM 模型参数 π_j^* 为定义在题目水平上的参数, 其下角标数字为题目编号, 代表特定题目所需属性均掌握的学生在该题目上正确作答的概率; 参数 r_{jk}^* 其下角标第一个数字为题目编号, 第二个数字为属性编号, 代表特定题目所需的某个属性未被掌握对题目正确作答概率的影响。

表7 不同知识状态的学生答对
不同题目概率的对数发生比(NIDO 模型)

KS		T4	T6	T7
		100	110	111
1	000	λ_{01}	$\lambda_{01} + \lambda_{02}$	$\lambda_{01} + \lambda_{02} + \lambda_{03}$
2	001	λ_{01}	$\lambda_{01} + \lambda_{02}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{13}$
3	010	λ_{01}	$\lambda_{01} + \lambda_{02} + \lambda_{12}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{12}$
4	011	λ_{01}	$\lambda_{01} + \lambda_{02} + \lambda_{12}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{12} + \lambda_{13}$
5	100	$\lambda_{01} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{11}$
6	101	$\lambda_{01} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{11} + \lambda_{13}$
7	110	$\lambda_{01} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{11} + \lambda_{12}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{11} + \lambda_{12}$
8	111	$\lambda_{01} + \lambda_{11}$	$\lambda_{01} + \lambda_{02} + \lambda_{11} + \lambda_{12}$	$\lambda_{01} + \lambda_{02} + \lambda_{03} + \lambda_{11} + \lambda_{12} + \lambda_{13}$

注:NIDO 模型中参数都用 λ 表示,其下角标的第一个数字为参数类型,0 表示基础率,1 表示斜率,第二个数字为属性编号。

表8 不同知识状态的学生答对
不同题目概率的对数发生比(C-RUM 模型)

KS		T4	T6	T7
		100	110	111
1	000	$-\pi_4^*$	$-\pi_6^*$	$-\pi_7^*$
2	001	$-\pi_4^*$	$-\pi_6^*$	$\tau_{73}^* - \pi_7^*$
3	010	$-\pi_4^*$	$\tau_{62}^* - \pi_6^*$	$\tau_{72}^* - \pi_7^*$
4	011	$-\pi_4^*$	$\tau_{62}^* - \pi_6^*$	$\tau_{72}^* + \tau_{73}^* - \pi_7^*$
5	100	$\tau_{41}^* - \pi_4^*$	$\tau_{61}^* - \pi_6^*$	$\tau_{71}^* - \pi_7^*$
6	101	$\tau_{41}^* - \pi_4^*$	$\tau_{61}^* - \pi_6^*$	$\tau_{71}^* + \tau_{73}^* - \pi_7^*$
7	110	$\tau_{41}^* - \pi_4^*$	$\tau_{61}^* + \tau_{62}^* - \pi_6^*$	$\tau_{71}^* + \tau_{72}^* - \pi_7^*$
8	111	$\tau_{41}^* - \pi_4^*$	$\tau_{61}^* + \tau_{62}^* - \pi_6^*$	$\tau_{71}^* + \tau_{72}^* + \tau_{73}^* - \pi_7^*$

注:模型参数中下角标的第一个数字代表题目编号,下角标的第二个数字(如果存在)代表属性编号。

是定义在题目和属性交叉水平上的参数,实质上是一个斜率,反映属性 k 的掌握体现在题目正确解答概率上的增益作用; π_j^* 是一个定义在题目水平上的截距,表示学生对题目 j 所需属性均未掌握的情况下凭猜测答对该题的程度,实为定义在题目水平上的基础率参数。表 8 列出了 C-RUM 模型下不同知识状态对应应在题目 j 上的作答概率。

四、认知诊断模型的内在关系梳理

认知诊断模型的基本思路是基于模型认知假设将有限的知识状态分成若干组,假定同一组内部不同知识状态的正确作答概率相同,不同组

之间作答概率各不同,同时假定学生对 q_j 向量中是否考查属性掌握不影响其正确作答的概率。例如对于考查 A1 和 A2 的题目 T6 而言,DINA 模型将 KS1 到 KS6 的正确作答概率归为一组,KS7 到 KS8 的归为一组,体现出非补偿的模型假设;DINO 模型则将 KS1 到 KS2 归为一组,KS3 到 KS8 归为一组,体现出补偿的模型假设;另外 4 个模型则一致将这 8 种知识状态分成四组,即 KS1 和 KS2 为第一组,KS3 和 KS4 为第二组,KS5 和 KS6 为第三组,KS7 和 KS8 为第四组,实际上是放松了模型假设,以提高模型和数据的拟合程度,缺点是参数结构更加复杂,带来参数估计上的困难。如果参照表 3 开发出一个有 7 个题目的认知诊断测验,若采用 DINA 和 DINO 模型均需估计 14 个题目参数,NIDA 和 NIDO 模型需要估计 6 个题目参数,而 C-RUM 和 R-RUM 模型需要估计 19 个题目参数。

表 4 所列的六种具体认知诊断模型适用于题目计分、属性掌握表征和题目对属性的考查都采用二值方式的情况,都以概率的形式表达了特定的知识状态在特定题目上正确作答的概率。这六种模型的区别如表 9 所示,主要体现在以下方面。①补偿模型和非补偿模型。依据属性之间的关系,DINA 模型、NIDA 模型和 R-RUM 属于非补偿模型,DINO 模型、NIDO 模型和 C-RUM 属于补偿模型。②参数类别、个数和定义的水平。DINA 模型和 DINO 模型的失误参数和猜测参数定义在题目水平上,NIDA 模型和 NIDO 模型的参数定义在属性水平上,但是参数的类别不同,NIDA 模型采用的依然是失误参数和猜测参数,NIDO 模型采用的是题目考查属性均未掌握的低水平学生作答反应概率的下限参数,和各个属性的掌握对正确作答概率的增益参数,R-RUM 和 C-RUM 的模型参数也更为复杂, π_j^* 均定义在题目水平上, τ_{jk}^* 均定义在属性和题目交叉水平上,但是同样符号在两个模型下的含义是相反的,R-RUM 下的 π_j^* 表示题目考查的属性均已掌握的高水平学生作答反应的上限参数,而在 C-RUM 下则表示题目考查属性均未掌握的低水平学生作

常见的认知诊断模型及其比较

答反应的下限参数 μ_{jk}^* 在 R-RUM 下表示考查属性未被掌握体现在正确作答概率上的惩罚作用,而在 C-RUM 中则体现为掌握了某个属性体现在正确作答概率上的增益作用。③连接函数,这六个模型中除 NIDO 模型和 C-RUM 采用的是 logit 函数形式,其他都采用了恒等的形式。这两种形式不仅体现出参数的量尺的不同,也体现了属性掌握和正确作答概率之间的关系是线性还是非线性的模型假设。这三方面的差异不仅导致了不同模型概率的分组差异,也造成了不同模型间横向比较和研究者之间交流的困难。

表 9 六种主要认知诊断模型设置的差异

模型	参数水平	参数类别	连接函数	题目参数个数	概率分组
DINA	题目	猜测/失误	恒等形式	$2 \times J$	2
NIDA	属性	猜测/失误	恒等形式	$2 \times K$	2^{K^*}
DINO	题目	猜测/失误	恒等形式	$2 \times J$	2
NIDO	属性	下限/增益	logit 形式	$2 \times K$	2^{K^*}
R-RUM	题目-属性	上限/惩罚	恒等形式	$J + \sum_{j=1}^J K_j^*$	2^{K^*}
C-RUM	题目-属性	下限/增益	logit 形式	$J + \sum_{j=1}^J K_j^*$	2^{K^*}

注:题目-属性代表参数定义在题目和属性的交叉环节,即为题目所考查的每个属性分别设定不同的参数。

五、评价与展望

认知诊断模型在认知诊断评估中起着至关重要的作用,它是连接学生的内在属性掌握模式和外在题目反应之间的桥梁。认知诊断评估在实施过程中往往需要综合考虑知识技能的表征方式、计分方式的客观限制和统计计算的可行性等诸多因素,来选择某一特定的认知诊断模型。然而,在教育测量实践中,将几个模型结合运用,从多维度、多方面分析学生的题目反应和测验的结构可能是更为合理的抉择。此外,认知诊断模型的参数估计颇为繁杂,需要专门的软件来实现,这可能是制约认知诊断评估推广的重要因素之一。

参考文献:

[1]Fu, J., Y. Li. Cognitively diagnostic psychometric models:An integrative review[P]. in Annual Meeting of the National Council on Measurement in Education. Chicago, IL, 2007.

[2]陈秋梅,张敏强.认知诊断模型发展及其应用方法述评[J].心理科学进展,2010,18(3):522~529.

[3][4]Junker, B.W., K. Sijtsma. Cognitive Assessment Models with Few Assumptions, and Connections with Nonparametric Item Response Theory [J]. Applied Psychological Measurement, 2001,25(3):258~272.

[5][8]Hartz, S.M.. A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality [D]. Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign, 2002.

[6]Templin, J.L., R.A. Henson, Measurement of psychological disorders using cognitive diagnosis models[J]. Psychological Methods, 2006,11(3):287~305.

[7]Templin, J.L.. CDM Users's guide [M]. Unpublished manuscript, 2006.

[9]涂冬波,蔡艳,戴海崎等.现代测量理论下四大认知诊断模型述评[J].心理学探新,2008,28(106):64~68.

责任编辑/王彩霞

