



认知诊断分析平台 (flexCDMs)



使 用 手 册

©博才智能测评技术研发团队

联系人：涂冬波 (tudongbo@aliyun.com)

目录

第一章 平台简介	1
一、认知诊断分析平台	1
二、平台主要功能	2
第二章 平台使用说明	4
一、Q 矩阵理论	4
（一）、Q 矩阵理论操作示例	4
（二）、Q 矩阵理论结果解读	6
二、参数估计	7
（一）、二级评分的参数估计操作示例	7
（二）、多级评分的参数估计操作示例	11
（三）、参数估计结果解读	15
三、模型拟合	16
（一）、模型拟合操作示例	16
（二）、模型拟合结果解读	18
四、测验质量分析	19
（一）、测验质量分析操作示例	19
（二）、测验质量分析结果解读	21
五、DIF 分析	22
（一）、DIF 分析操作示例	22
（二）、DIF 分析结果解读	24
六、Q 矩阵修正与估计	25

(一)、Q 矩阵修正操作示例	25
(二)、Q 矩阵估计操作示例	28
(三)、Q 矩阵修正和估计的结果解读	33
七、图形输出	34
(一)、图形输出操作示例.....	34
(二)、图形输出结果解读.....	36
八、数据服务	38
(一)、实测数据	38
(二)、模拟数据	39
第三章 参考文献.....	42

第一章 平台简介

一、认知诊断分析平台

1、平台宗旨

致力于为广大实际应用者及爱好者提供全面、可靠、简单、方便的认知诊断（Cognitive Diagnosis）数据分析服务，以推动认知诊断在我国的进一步发展及推广。

2、平台主界面



3、平台网址与浏览器要求

平台网址为 www.psychometrics-studio.cn，本平台推荐使用谷歌浏览器或火狐浏览器访问，不支持手机浏览。



4、联系人

涂冬波 (tudongbo@aliyun.com)

二、平台主要功能

本平台主要包括八大功能模块（详见图 1），各功能模块的具体功能描述详见表 1。

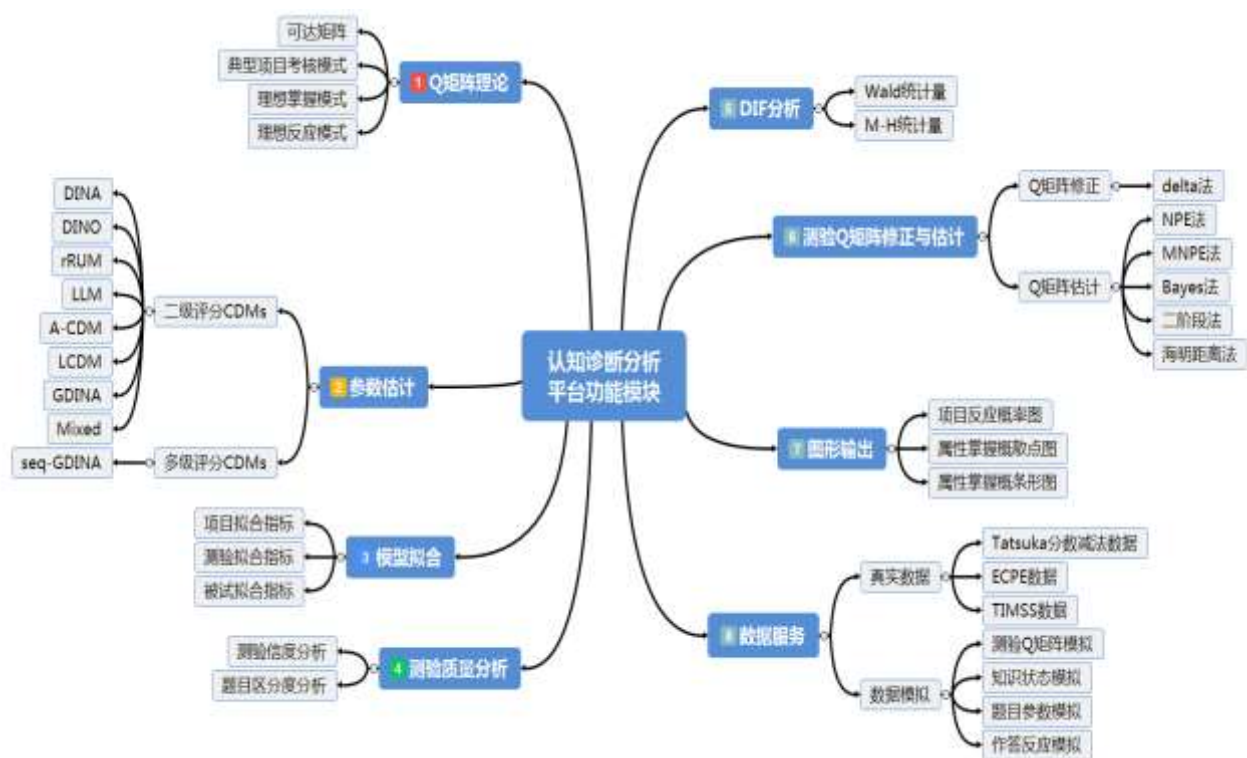


图 1 flexCDMs 功能模块

表 1 flexCDMs 具体功能描述

功能模块序号	名称	具体功能描述
1	Q 矩阵理论	Tatsuoka 的 Q 矩阵理论：包括根据邻接矩阵（A 阵）生成可达矩阵（R 阵）、理想测量模式以及理想掌握模式。
2	参数估计	常用认知诊断模型（CDMs）参数估计实现：包括二级评分 CDM（如 G-DINA, LCDM, A-CDM, LLM, rRUM, DINO 和 DINA）和多级评分 CDM（seq-GDINA）。
3	模型拟合	检验数据与模型的拟合度：包括项目拟合（绝对水平拟合）检验、测验拟合（相对水平拟合）检验及被试拟合检验。
4	测验质量分析	测验质量是判断一个测验性能的重要标准之一，本模块的主要功能包括基于 CTT 及认知诊断的测验信度分析，以及基于认知诊断的项目区分度分析。
5	DIF 分析	基于认知诊断框架下检测项目功能差异（DIF）：具体方法包括 Wald 法以及 M-H 法。
6	测验 Q 矩阵修正与估计	Q 矩阵辅助界定：包括 Q 矩阵修正以及 Q 矩阵估计两大功能。本模块中 Q 矩阵修正的方法为 delta 法；；Q 矩阵估计的方法有 NPE 法、MNPE 法、Bayes 法以及二阶段法以及海明距离法。
7	图形输出	数据分析结果的图形输出：目前可提供的图形包括 ICC（项目特征曲线）、被试测验总分与每个属性后验掌握概率散点图、以及被试测验总分分组与属性后验掌握概率条形图。
8	数据服务	包括两个部分：数据模拟功能（Q 矩阵、属性掌握模式、项目参数、作答反应模式）及实测数据提供。可使用的实测数据包括 ECPE 数据、TIMSS 数据以及 Tatsuoka 分数减法数据。

第二章 平台使用说明

一、Q 矩阵理论

(一)、Q 矩阵理论操作示例

1、在功能模块栏中选择“Q 矩阵理论”



2、进入 Q 矩阵理论界面，点击 Browse 键上传 A 阵。上传的数据要求是 CSV 文件（Excel 文件可以另存为 CSV 文件）。



3、A 矩阵文件示例

	A	B	C	D	E	F	G	H	I	J	K
1		A1	A2	A3	A4						
2	A1		0	1	0	1					
3	A2		0	0	1	0					
4	A3		0	0	0	0					
5	A4		0	0	0	0					
6											
7											
8											
9											
10											
11											
12											
13											

A 矩阵（Adjacency Matrix）反映的是认知属性之间的直接关系（既不含间接关系和自身关系）的矩阵。它由 K 行 K 列（K 值属性个数）的 0-1 矩阵组成，如果认知属性之间存在直接关系，则在邻接矩阵（A 矩阵）相应元素中用“1”表示，否则用“0”表示。

4、计算并提取结果



点击 “开始计算” 进行计算

计算结果 下方会显示计算结果，根据需要可点击选择 “R矩阵”、“理想测量模式”、“理想掌握模式” 以及 “理想反应模式”

R矩阵

理想测量模式

理想掌握模式

理想反应模式

Show 5 entries

	A1	A2	A3	A4
A1	1	1	1	1
A2	0	1	1	0
A3	0	0	1	0
A4	0	0	0	1

Showing 1 to 4 of 4 entries

Previous

1

Next

5、结果导出

结果导出 点击需要保存的结果下载保存

保存R矩阵

保存理想测量模式

保存理想掌握模式

保存理想反应模式

（二）、Q 矩阵理论结果解读

结果类型	结果解读
R 矩阵	R 矩阵 (Reachability Matrix), 也称可达矩阵。反映认知属性间直接关系、间接关系、和自身关系的矩阵。由 K 行 K 列 (K 指属性个数) 的 0-1 矩阵组成, 如果认知属性之间存在直接关系、间接关系或自身关系中的任何一种, 则在 R 矩阵 (可达矩阵) 相应的元素中用 “1” 表示, 否则用 “0” 表示。
理想测量模式	理想测量模式, 也称典型项目考核模式。是指根据属性间的层级关系, 确定所有合逻辑的测验项目考核模式。它是一个 K 列 0-1 矩阵, “1” 代表题目测量了该属性, “0” 代表未测量。
理想掌握模式	理想掌握模式指根据属性间的层级关系, 符合逻辑的掌握模式。理想掌握模式通常也被称为知识状态 (knowledge states) 或认知结构。他与理想测量模式类似, 对于测验项目而言, 有属性是否被测量或考核之分; 对于被试, 也有属性是否掌握之分。理想掌握模式的获取与理想测量模式的获取原理一致, 但是比理想测量模式多一种, 即全为 0 的模式。
理想反应模式	理想反应模式是指被试在不存在任何失误和猜测等误差条件下对项目的作答反应情况。既若被试掌握了项目考核的所有属性则被试答对该题, 若被试至少有一个项目考核属性未掌握, 则被试答错该项目。例如, 掌握模式为 (1100) 的被试对项目考核模式为 (1000) 和 (1101) 的两道题的理想反应为 (10), 既能答对第一题, 但答错第二题。如果测验被试数为 I , 项目数为 J , 则理想反应模式是一个 I 行 J 列的 0-1 矩阵, 0 代表答对, 1 代表答错。

二、参数估计

（一）、二级评分的参数估计操作示例

1、在功能模块栏中选择“参数估计”选项，在下拉菜单中选择“二级评分 CDMs”选项。



2、进入参数估计界面，上传数据文件。上传的数据要求是 CSV 文件。



3、上传数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	0				
5	ID4	1	1	1	1	1	1	1				
6	ID5	1	1	1	1	1	1	1				
7	ID6	1	1	1	1	1	1	1				
8	ID7	1	1	1	1	1	1	1				
9	ID8	0	1	1	1	1	1	0				
10	ID9	1	1	1	1	1	1	1				
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1				
14	ID13	1	1	1	1	1	1	1				
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1				
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

4、数据预览

(3) 数据预览:

作答数据

Q矩阵

Click "作答数据" or "Q矩阵" to preview corresponding data

Show 5 entries

选择每页显示的行数

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
ID1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0
ID2	1	1	1	0	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	1
ID3	0	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1
ID4	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	1
ID5	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	1	0	0

Showing 1 to 5 of 2,019 entries

Previous

1

2

3

4

5

...

404

Next

可选择预览的数据页码

5、计算与保存结果

选择模型进行参数估计

(4) 选择估计模型:

☒ DINA ☐ DINO ☐ rRUM ☐ LLM ☐ ACDM ☐ GDM ☐ LCDM ☐ GDINA ☐ Mixed

✓ 开始估计

点击“开始估计”进行计算

5.1、被试 alpha 估计结果:

计算结果

被试alpha

试题参数

属性反应概率

点击“被试alpha”可以查看被试属性掌握模式的估计结果

选择查看 被试参数

点击“选择查看被试参数”下面的下拉菜单，可以选择查看不同估计方法得到的被试参数估计结果，其中“属性掌握概率”为每位被试掌握各个属性的概率。

		A2	A3	A4
MLE		0	0	0
MAP		1	0	1
EAP		1	0	1
属性掌握概率		0	1	0
ID2	0	0	0	0
ID3	0	1	0	1
ID4	0	0	1	0
ID5	0	0	0	0

Showing 1 to 5 of 2,019 entries

Previous 1 2 3 4 5 ... 404 Next

5.2、试题参数估计结果:

计算结果

被试alpha

试题参数

属性反应概率

点击“试题参数”可以查看试题参数估计结果

选择查看 项目参数

在“选择查看项目参数”下面的下拉菜单中，可以选择需要查看的项目参数类型。其中“选用的混合模型”一项只有在“选择估计模型”中选择“Mixed”时才可查看。

	P(0)	1-P(1)
Item 1	0.5178	0.1567
Item 2	0.8224	0.0119
Item 3	0.8148	0.0315
Item 4	0.9448	0.005
Item 5	0.2603	0.2451

Showing 1 to 5 of 20 entries

Previous 1 2 3 4 Next

5.3、属性反应概率结果:

计算结果

被试alpha 试题参数 属性反应概率 选择“属性反应概率”选项

Show: 5 entries

	H	I	D	M	S	W	U	Q	Y	Q	H1	H2	H3	H4	H5	H6	H7	H8	H9	U0
0000	0.538	0.7464	0.7889	0.9431	0.2914	0.4722	0.8366	0.4499	0.808	0.1381	0.7303	0.5687	0.8643	0.8538	0.3438	0.2601	0.1908	0.3095	0.2728	0.2586
1000	0.538	0.8579	0.7889	0.9431	0.2914	0.4722	0.8366	0.4499	0.808	0.1381	0.7303	0.5687	0.8643	0.8538	0.3438	0.2601	0.1908	0.3095	0.2728	0.2586
0100	0.538	0.7464	0.7889	0.9431	0.2914	0.4722	0.8366	0.4499	0.808	0.1381	0.7303	0.5687	0.8643	0.8538	0.3438	0.2601	0.1908	0.3095	0.2728	0.2586
0910	0.538	0.7464	0.7889	0.9950	0.2914	0.4722	0.9819	0.4499	0.9918	0.1381	0.8872	0.5687	0.8643	0.8538	0.3438	0.2601	0.1908	0.3095	0.2728	0.8483
0901	0.538	0.7464	0.9667	0.9431	0.2914	0.4722	0.8366	0.4499	0.808	0.1381	0.7303	0.5687	0.9963	0.8538	0.3438	0.2601	0.1908	0.3095	0.2728	0.2586

Showing 1 to 5 of 16 entries

Previous 1 2 3 4 Next

该结果表示的是各种理想掌握模式在各个题目上的答对概率。每一行代表每一种理想属性掌握模式，每一列代表各个题目。

5.3、结果保存

结果导出

选择需要保存的结果

保存被试alpha 保存试题参数 属性反应概率

这里需要注意的是，保存的结果与选择查看的结果一致。例如：在选择查看被试参数中查看的是 EAP 算法估计出来的被试 alpha 参数，则在结果导出中，点击保存被试 alpha 后保存的也是 EAP 算法估计出来的被试 alpha 参数。

(二)、多级评分的参数估计操作示例

1、在功能模块栏中选择“参数估计”选项，在下拉菜单中选择“多级评分 CDMs”选项。



2、进入参数估计界面，上传数据文件。



3、上传数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	
1		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
2	ID1	0	2	0	2	2	0	0	1	0	0	
3	ID2	2	0	0	0	0	1	1	1	0	0	
4	ID3	0	0	2	2	0	0	2	1	0	0	
5	ID4	0	0	0	0	0	0	0	0	0	0	
6	ID5	0	2	1	1	1	0	0	1	2	0	
7	ID6	0	0	2	2	0	0	1	1	1	0	
8	ID7	0	0	0	0	0	0	1	1	0	0	
9	ID8	1	0	2	2	1	2	2	2	2	0	
10	ID9	1	0	2	1	0	1	0	0	0	0	
11	ID10	0	0	2	2	2	1	1	2	0	0	
12	ID11	1	1	2	1	1	1	0	0	0	0	
13	ID12	2	2	0	0	0	2	1	1	0	1	
14	ID13	1	1	0	0	1	0	1	0	0	0	
15	ID14	0	1	0	0	2	0	1	0	0	1	
16	ID15	1	0	2	2	0	1	1	2	1	0	
17	ID16	0	0	1	1	0	0	0	0	1	0	
18	ID17	0	1	0	0	1	0	0	0	0	0	
19	ID18	1	2	2	0	0	1	0	2	2	0	
20	ID19	2	1	0	0	1	1	0	0	0	0	

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Item	Cat	A1	A2	A3	A4	A5					
2	1	1	1	1	0	0	0	0					
3	2	1	2	0	1	0	0	0					
4	3	2	1	0	0	1	0	0					
5	4	2	2	0	0	0	1	0					
6	5	3	1	0	0	0	0	1					
7	6	3	2	1	0	0	0	0					
8	7	4	1	0	0	0	0	1					
9	8	4	2	0	0	0	1	0					
10	9	5	1	0	0	1	0	0					
11	10	5	2	0	1	0	0	0					
12	11	6	1	1	0	0	0	0					
13	12	6	2	0	1	1	0	0					
14	13	7	1	0	1	0	0	0					
15	14	7	2	0	0	1	1	0					
16	15	8	1	0	0	0	1	0					
17	16	8	2	1	0	0	0	1					
18	17	9	1	0	0	0	1	1					
19	18	9	2	0	0	1	0	0					
20	19	10	1	0	1	1	0	0					

Q矩阵数据格式示例

多级评分的Q矩阵文件相较二级评分的Q矩阵多出了序列号以及Cat列,Cat为category的简写,意为题目每一步所测量的属性模式。例如:示例数据中第一题第一步(得1分)测量的属性模式Cat1为(10000)即第一题第一步只测量了第一个属性,第二步(得2分)Cat2所需要的属性掌握模式为(01000)即第一题第二步只测量了第二个属性,以此类推。关于seq-GDINA模型的Q矩阵要求,请读者详见文献Ma和de la Toore (2016)。

4、数据预览

(3) 数据预览:

[作答数据](#)
[Q矩阵](#)

点击“作答数据”或“Q矩阵”可预览相应数据

选择每页显示的行数

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
ID1	0	2	0	2	2	0	0	1	0	0	0	0	3	0	2	0	0	1	1	1
ID2	2	0	0	0	0	1	1	1	0	0	1	1	0	2	0	1	0	0	1	0
ID3	0	0	2	2	0	0	2	1	0	0	0	0	2	0	2	0	0	0	1	1
ID4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
ID5	0	2	1	1	1	0	0	1	2	0	2	0	3	0	1	0	0	1	0	1

Showing 1 to 5 of 2,000 entries

Previous 1 2 3 4 5 ... 400 Next

可选择预览的数据页码

5、计算与保存结果

(4) 选择估计模型:

☒ seq-GDINA

选择模型进行参数估计

点击“开始估计”进行计算

5.1、被试 alpha 估计结果:

[被试alpha](#)
[试题参数](#)
[属性反应概率](#)

点击“被试alpha”可以查看被试属性掌握模式的估计结果

选择查看被试参数

MLE|

 MLE

 MAP

 EAP

 属性掌握概率

点击“选择查看被试参数”下面的下来菜单，可以选择查看不同估计方法得到的被试参数估计结果，其中“属性掌握概率”为每位被试掌握各个属性的概率。

	A2	A3	A4	A5
ID1	0	1	1	1
ID2	1	0	0	0
ID3	0	0	1	1
ID4	0	0	0	0
ID5	0	1	1	1

Showing 1 to 5 of 2,000 entries

Previous 1 2 3 4 5 ... 400 Next

选择查看的页码

5.2、试题参数估计结果:

被试alpha

试题参数

属性反应概率

点击“试题参数”查看试题参数估计结果

选择查看项目参数

gs

gs

delta

在“选择查看项目参数”下面的下拉菜单中，可以选择需要查看的项目参数类型。

	P(0)	1-P(1)
Item 1	0.099	0.1087
Item 2	0.0941	0.0995
Item 3	0.1048	0.0977
Item 4	0.0988	0.1138
Item 5	0.0987	0.0945

Showing 1 to 5 of 39 entries

Previous

1

2

3

4

5

...

8

Next

选择查看的页码

5.3、属性反应概率结果：

被试alpha

试题参数

属性反应概率

点击“属性反应概率”选项

Show 5 entries

选择每页显示的行数

Item 1 Cat 1	Item 1 Cat 2	Item 2 Cat 1	Item 2 Cat 2	Item 3 Cat 1	Item 3 Cat 2	Item 4 Cat 1	Item 4 Cat 2	Item 5 Cat 1	Item 5 Cat 2	Item 6 Cat 1	Item 6 Cat 2	Item 7 Cat 1	Item 7 Cat 2	Item 8 Cat 1	Item 8 Cat 2
00000	0.0897	0.0093	0.0945	0.0104	0.0876	0.0111	0.1022	0.0097	0.0924	0.011	0.1043	0.0119	0.0999	0.011	0.0963
10000	0.8074	0.0839	0.0945	0.0104	0.0105	0.0882	0.1022	0.0097	0.0924	0.011	0.7977	0.0909	0.0999	0.011	0.1003
01000	0.0099	0.0891	0.0945	0.0104	0.0876	0.0111	0.1022	0.0097	0.0086	0.0948	0.1027	0.0135	0.8107	0.0891	0.0963
00100	0.0897	0.0093	0.8132	0.0892	0.0876	0.0111	0.1022	0.0097	0.8006	0.095	0.1074	0.0088	0.1002	0.0106	0.0963
00010	0.0897	0.0093	0.0119	0.0929	0.0876	0.0111	0.0121	0.0998	0.0924	0.011	0.1043	0.0119	0.0982	0.0127	0.8213

Showing 1 to 5 of 32 entries

Previous

1

2

3

4

5

6

7

Next

选择想要查看的页码

该结果表示的是各种理想掌握模式在各个题目上的答对概率。每一行代表每一种理想属性掌握模式，每一列代表各个题目。

5.4、结果保存

结果导出

选择需要保存的结果

保存被试alpha

保存试题参数

属性反应概率

这里需要注意的是，保存的结果与选择查看的结果一致。例如：在选择查看被试参数中查看的是 EAP 算法估计出来的被试 alpha 参数，则在结果导出中，点击保存被试 alpha 后保存的也是 EAP 算法估计出来的被试 alpha 参数。

(三)、参数估计结果解读

1、被试 alpha 结果解读

结果名称	结果解读
MLE	极大似然估计 (Maximum Likelihood Estimation, MLE), 既为被试项目反应概率的联合似然函数极大值点对应的属性掌握模式
MAP	极大后验估计 (Maximum a Posteriori Estimation, MAPE)。在统计学中, 极大后验估计是在被试样本数据已知的情形下, 获得对未知参数的点估计。它与基于经典统计理论的极大似然估计方法非常相似, 只是极大后验估计这种方法在构建似然函数时, 要将未知参数的先验分布融合到似然函数式中。所以, 极大后验估计可以看做是经过正则化 (regularization) 之后的极大似然估计。该方法简单易用, 且估计精度较高。
EAP	贝叶斯期望后验估计 (Expected a Posterior Estimation, EAP)。该方法的特点是使用后验分布来总结数据、得到推论。贝叶斯方法试图算出后验均值或者中值, 而不是寻找一个确定极值-众数。期望后验估计具有简单、高效、稳定的特点, 是能力参数估计方法中的较佳选择。
属性掌握概率	每位被试掌握各个属性的后验概率, 每一行为每一位被试, 每一列为各个属性。

2、试题参数结果解读

结果名称	结果解读
gs	类似于 DINA 模型的 gs 参数, 第一列 P (0) 表示未掌握该项目任一测量属性的被试而答对项目的概率(猜测概率 g), 第二列 1-P (1) 表示掌握了该项目测量所有属性的被试而答错项目的概率 (失误参数 s), 每一行代表各个题目
delta	各个模型与 GDINA 模型相对应的参数, 具体内容请参见参考文献。
选用的混合模型 (仅限二级评分 CDMs)	使用 Wald 法为每道题目选出的 GDINA 模型族中最适合的模型。具体内容请参见参考文献。

3、属性反应概率

结果名称	结果解读
属性反应概率 (二级评分 CDMs)	各种理想掌握模式在各个题目上的正确作答概率
属性反应概率 (多级评分 CDMs)	各种理想掌握模式在各个题目及分步 (Cat) 上的正确作答概率

三、模型拟合

(一)、模型拟合操作示例

1、选择拟合指标类型。



2、点击 Browse 键分别上传作答数据和 Q 矩阵，上传的数据要求是 CSV 文件。作答数据文件第一行为题目序号，第一列为被试序号，一行代表一个被试，一列代表一个题目；Q 矩阵数据文件第一行为属性序号，第一列为题目序号，一行代表一个题目，一列代表一个属性。

⚙ 参数设定

(1) 上传作答数据：

Browse...

No file selected

点击Browse键上传作答数据文件

(2) 上传Q矩阵：

Browse...

No file selected

点击Browse键上传Q矩阵文件

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	0				
5	ID4	1	1	1	1	1	1	1				
6	ID5	1	1	1	1	1	1	1				
7	ID6	1	1	1	1	1	1	1				
8	ID7	1	1	1	1	1	1	1				
9	ID8	0	1	1	1	1	1	0				
10	ID9	1	1	1	1	1	1	1				
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1				
14	ID13	1	1	1	1	1	1	1				
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1				
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

3、选择要与哪个模型进行拟合度检验，并点击“开始估计”按钮进行计算。

(4) 选择估计模型：

☒ DINA ☐ DINO ☐ rRUM ☐ LLM ☐ ACDM ☐ GDM ☐ LCDM ☐ GDINA ☐ Mixed

选择模型

✓ 开始估计

点击“开始估计”按钮开始计算

4、在线查看结果或导出结果。

计算结果

题目拟合(Maximum Statistics, Chen et al, 2013)

题目拟合(RMSEA)

题目拟合(SX2)

在线查看结果

结果导出

离线保存结果

保存Maximum Statistics

保存RMSEA

保存SX2

(二)、模型拟合结果解读

具体指标		评价标准
项目拟合指标	Maximum Statistics	adj. pvalue.max[z. r], adj. pvalue.max[z. logOR]> 0.05 表示拟合良好
	RMSEA	RMSEA 的值越接近 0, 说明项目拟合越好。在 Oliveri 和 von Davier (2011) 的研究中, RMSEA 的临界值被设定为 0.1, RMSEA>0.1 时说明项目拟合较差。
	SX2	P<0.05, 说明项目拟合较差
测验拟合指标	偏差	相对指标, 其值越小表明模型拟合越好
	AIC	相对指标, 其值越小表明模型拟合越好
	BIC	相对指标, 其值越小表明模型拟合越好
被试拟合指标	Lz	沿用 IRT 的传统, 其值小于-2 表示不拟合 ($Lz < -2$)

四、测验质量分析

（一）、测验质量分析操作示例

1、选择分析测验信度或题目区分度。



2、点击 Browse 键分别上传作答数据和 Q 矩阵，上传的数据要求是 CSV 文件。作答数据文件第一行为题目序号，第一列为被试序号，一行代表一个被试，一列代表一个题目；Q 矩阵文件第一行为认知属性序号，第一列为题目序号，一行代表一个题目，一列代表一个属性。

参数设定

(1) 上传作答数据：

Browse...

No file selected

点击Browse键上传作答数据文件

(2) 上传Q矩阵：

Browse...

No file selected

点击Browse键上传Q矩阵文件

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	1	0			
5	ID4	1	1	1	1	1	1	1	1			
6	ID5	1	1	1	1	1	1	1	1			
7	ID6	1	1	1	1	1	1	1	1			
8	ID7	1	1	1	1	1	1	1	1			
9	ID8	0	1	1	1	1	1	1	0			
10	ID9	1	1	1	1	1	1	1	1			
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1	1			
14	ID13	1	1	1	1	1	1	1	1			
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1	1			
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

3、选择采用哪个模型来分析测验信度或题目区分度。

(4) 选择估计模型：

☒ DINA
 ☐ DINO
 ☐ rRUM
 ☐ LLM
 ☐ ACDM
 ☐ GDM
 ☐ LCDM
 ☐ GDINA
 ☐ Mixed
 选择模型

点击“开始估计”按钮开始计算

4、在线查看结果或导出结果。

⚙️ 计算结果

在线查看结果

⚙️ 结果导出

离线下载结果

（二）、测验质量分析结果解读

1、测验信度结果解读

本平台共提供三种测验信度的结果，它们分别是基于 CTT 的信度，Templin 和 Bradshaw（2013）的属性重测一致性指标和 Cui 等人（2012）开发的分类一致性信度指标。有关 CTT 的信度，读者可参考相关的心理测量书籍，本手册具体介绍后两种基于认知诊断的信度。

Templin 和 Bradshaw(2013)提出的属性重测一致性指标与 CTT 中标准参照测验中决策一致性的 Subkoviak 方法类似，假设被试的属性掌握概率不变，则可求得被试在两次重测时属性掌握的四格相关，以四格相关作为认知诊断测验属性信度的指标。Cui、Girel 和 Chang(2012)的分类一致性信度指标则是沿用 Lee 等人(2002)在 CTT 和 IRT 中计算分类一致性信度的思路而提出的。

2、题目区分度结果解读

区分度指标反应题目对不同属性掌握模式的被试的区分能力，本平台中题目区分度等于掌握项目测量所有属性被试答对的概率[即 $P(1)$]减去未掌握项目测量的任一属性被试答对的概率[即 $P(0)$]，即

$$\text{Discrimination} = P(1) - P(0)$$

题目区分度的取值范围在-1 到 1 之间，其值越大表明题目质量越好，根据 CTT 标准（如 CTT 中的鉴别指数区分度），一般来讲区分度大小 0.4 则较为优秀。

五、DIF 分析

（一）、DIF 分析操作示例

1. 选择 DIF 功能模块，该模块提供了两种 DIF 分析方法：Wald 法以及 M-H 法，这里以 Wald 法为例进行演示。



2. 上传作答数据、Q 矩阵以及分组数据

参数设定

(1) 上传作答数据:

Browse...

MELAB_data.csv

Upload complete

点击 "Browse" 上传作答数据文件

(2) 上传Q矩阵:

Browse...

MELAB_Q.csv

Upload complete

点击" Browse "上传Q矩阵文件

(3) 上传分组数据:

Browse...

MELAB_group.csv

Upload complete

点击" Browse "上传分组数据

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	0				
5	ID4	1	1	1	1	1	1	1				
6	ID5	1	1	1	1	1	1	1				
7	ID6	1	1	1	1	1	1	1				
8	ID7	1	1	1	1	1	1	1				
9	ID8	0	1	1	1	1	1	0				
10	ID9	1	1	1	1	1	1	1				
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1				
14	ID13	1	1	1	1	1	1	1				
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1				
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

	A	B	C	D	E	F	G	H	I	J	K
1		group									
2	ID1	1									
3	ID2	1									
4	ID3	1									
5	ID4	1									
6	ID5	1									
7	ID6	1									
8	ID7	1									
9	ID8	1									
10	ID9	1									
11	ID10	1									
12	ID11	1									
13	ID12	1									

分组数据格式示例

3. 预览上传的数据

(4) 数据预览:

作答数据

Q矩阵

分组数据

选择想要预览的数据文件

Show 5 entries 选择每页显示的行数

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
ID1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0
ID2	1	1	1	0	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	1
ID3	0	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1
ID4	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	1
ID5	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	1	0	0

Showing 1 to 5 of 2019 entries

Previous

1

2

3

4

5

...

404

Next

选择想要查看的页数

4. 计算与结果保存

开始估计

点击“开始估计”进行计算

DIF检验指标

Show 5 entries

选择每页显示的行数

	Wald stat.	df	p.value
Item 1	7.3802	4	0.1171
Item 2	0.0819	2	0.9599
Item 3	46.6306	2	0
Item 4	27.3584	2	0
Item 5	23.6083	4	0.0001

Showing 1 to 5 of 20 entries

选择所要查看的页数

Previous 1 2 3 4 Next

结果导出

保存DIF检验指标

点击“保存DIF检验结果”进行保存

M-H 检验方法的操作与 Wald 法一致，只是增加了一个模型选择的部分，使用者只需要选择拟合数据所用的认知诊断模型即可完成 DIF 检验，具体操作不再赘述。

(二)、DIF 分析结果解读

具体指标	评价标准
Wald 法	p. value 小于 0.05 则可以认为该题存在 DIF
M-H 法	ETS 公司根据 MH 方法计算的结果，把项目分成三种等级： 等级 A: delat-mh 的绝对值小于 1，或者与 0 没有显著差异($p>0.05$)。这类项目被视为具有很小的 DIF 或者无 DIF，可以用于测验。 等级 C: delat-mh 的绝对值大于 1.5，并且显著大于 1 ($P<0.05$)。这类项目被视为具有严重的 DIF，除非专家认为项目对测验至关重要，否则该类项目应被删除。 等级 B: 除了 A 和 C 的 delat-mh 值，这类项目只有当 A 类项目不能满足测验内容要求时，才能用于测验。

注：本平台 M-H 检验方法结果中的 delta_amh 为 delat-mh 的绝对值，mh_ka 为 delat-mh 卡方检验的显著性水平（P 值）。

六、Q 矩阵修正与估计

（一）、Q 矩阵修正操作示例

1、在功能模块栏中点击“Q 矩阵修正与估计”选项，下拉菜单中会出现“Q 矩阵修正”以及“Q 矩阵估计”两个菜单，选择“Q 矩阵修正”



2、上传作答数据以及原始 Q 矩阵

参数设定

(1) 上传作答数据：

Browse...

No file selected

点击Browse键上传作答数据文件

(2) 上传Q矩阵：

Browse...

No file selected

点击Browse键上传Q矩阵文件

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	0				
5	ID4	1	1	1	1	1	1	1				
6	ID5	1	1	1	1	1	1	1				
7	ID6	1	1	1	1	1	1	1				
8	ID7	1	1	1	1	1	1	1				
9	ID8	0	1	1	1	1	1	0				
10	ID9	1	1	1	1	1	1	1				
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1				
14	ID13	1	1	1	1	1	1	1				
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1				
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

3、预览数据

(3) 数据预览:

☒ 作答数据 ☐ Q矩阵

点击“作答数据”或“Q矩阵”可预览相应数据

Show 5 entries

选择每页显示的行数

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
ID1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0
ID2	1	1	1	0	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	1
ID3	0	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1
ID4	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	1
ID5	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	1	0	0

Showing 1 to 5 of 2,019 entries

Previous 1 2 3 4 5 ... 404 Next

可选择预览的数据页码

4、选择模型并开始计算

(4) 选择估计模型:

☒ DINA ☐ DINO ☐ rRUM ☐ LLM ☐ ACDM ☒ GDINA ☐ Mixed

选择拟合数据所用的模型

☒ 开始估计

点击“开始估计”进行计算

修正的Q矩阵

在下方查看估计结果

Show 5 entries

	A1	A2	A3	A4
Item1	1	0	0	0#
Item2	1	0	0	0
Item3	0	0	0	1
Item4	0	1#	1	1#
Item5	1	0	0	1

Showing 1 to 5 of 20 entries

Previous 1 2 3 4 Next

结果显示中，带有“#”的位置表明被修正的位置，“0#”表示原始Q矩阵此处为1，修正后为0；“1#”表示原始Q矩阵此处为0，修正后为1。

5、保存结果

⚙ 结果导出

⬇ 保存修正的Q矩阵

点击“保存修正的Q矩阵”保存计算结果

(二)、Q 矩阵估计操作示例

1. 选择“Q 矩阵估计”选项



2. 上传数据并预览

参数设定

(1) 上传作答数据：

Browse... No file selected

点击Browse键上传作答数据文件

(2) 上传Q矩阵：

Browse... No file selected

点击Browse键上传Q矩阵文件

(3) 数据预览：



点击“作答数据”或“Q矩阵”可预览相应数据

选择每页显示的行数

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20
ID1	0	1	1	1	1	1	1	0	1	0	0	1	1	1	0	0	0	1	0	0
ID2	1	1	1	0	1	0	1	0	1	0	0	1	1	1	0	0	1	0	1	1
ID3	0	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	0	0	1
ID4	1	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	1
ID5	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	0	0	1	0	0

Showing 1 to 5 of 2,019 entries

Previous 1 2 3 4 5 ... 404 Next

可选择预览的数据页码

3. 数据格式说明

3.1 海明距离法 Q 矩阵数据格式

海明距离法要求 Q 矩阵文件包含部分题目的 Q 矩阵信息，使用者需尽量保证 Q 矩阵的正确性与丰富性，具体信息请参见参考文献（汪大勋&涂冬波，2017）。

	A	B	C	D	E	F	G	H	I	J	K	L
1	itemId	skill1	skill2	skill3	skill4	skill5						
2	item1	1	0	0	0	0						
3	item16	0	0	0	0	1						
4	item29	1	0	1	1	1						
5	item30	0	1	1	1	1						
6	item31	1	1	1	1	1						
7												
8												
9												
10												
11												
12												
13												

海明距离法Q矩阵文件格式示例

3.2 贝叶斯方法 Q 矩阵数据格式

	A	B	C	D	E	F	G	H	I	J	K
1	itemId	skill1	skill2	skill3	skill4	skill5					
2	item1	1	0	0	0	0					
3	item2	0	1	0	0	0					
4	item3	NA	NA	0	0	0					
5	item4	0	0	1	0	0					
6	item5	1	0	1	0	0					
7	item6	NA	1	1	0	0					
8	item7	1	1	1	0	0					
9	item8	0	0	0	1	0					
10	item9	1	0	0	1	0					
11	item10	0	1	0	1	0					
12	item11	1	1	0	1	0					

贝叶斯方法Q矩阵示例

数据文件中“NA”表示不确定的属性标定。

3.3、其他方法的 Q 矩阵数据格式

	A	B	C	D	E	F	G	H	I	J	K
1	itemId	skill1	skill2	skill3	skill4	skill5					
2	item1	1	0	0	0	0					
3	item2	0	1	0	0	0					
4	item3	1	1	0	0	0					
5	item4	0	0	1	0	0					
6	item5	1	0	1	0	0					
7	item6	0	1	1	0	0					
8	item7	1	1	1	0	0					
9	item8	0	0	0	1	0					
10	item9	1	0	0	1	0					
11	item10	0	1	0	1	0					
12	item11	1	1	0	1	0					

其他方法 (NEP、MNEP、TwoStage法)
数据文件

3.4、作答数据文件格式

所有方法的作答数据格式均相同，具体如下：

	A	B	C	D	E	F	G	H	I	J	K	L
1	examineId	item1	item2	item3	item4	item5	item6	item7	item8	item9	item10	item11
2	ex1	1	1	1	1	1	0	1	0	0	0	0
3	ex2	0	1	1	0	0	1	1	0	0	1	0
4	ex3	0	1	1	0	1	0	1	1	1	0	1
5	ex4	0	0	0	1	1	0	0	0	1	0	0
6	ex5	1	0	1	1	1	0	0	0	0	1	1
7	ex6	1	0	0	1	1	0	0	0	0	0	0
8	ex7	1	0	0	0	1	0	1	1	0	0	0
9	ex8	1	1	1	1	1	0	1	0	1	0	0
10	ex9	0	1	0	0	1	1	1	0	1	0	0
11	ex10	0	0	0	0	1	1	1	0	1	0	0
12	ex11	0	1	0	1	1	0	1	1	0	0	0
13	ex12	0	1	0	0	1	0	1	0	1	1	1
14	ex13	1	1	0	1	1	0	1	0	0	0	0
15	ex14	1	1	1	0	1	0	0	1	0	1	0

作答数据文件格式示例

4. 预览数据

(3) 数据预览:

点击“作答数据”或“Q矩阵”可预览相应数据

Show 5 entries 选择每页显示的行数

Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Item15
ex1	1	1	1	1	1	0	1	0	0	0	0	0	0	0
ex2	0	1	1	0	0	1	1	0	0	1	0	0	0	0
ex3	0	1	1	0	1	0	1	1	1	0	1	0	0	1
ex4	0	0	0	1	1	0	0	0	1	0	0	0	0	1
ex5	1	0	1	1	1	0	0	0	0	1	1	1	0	1

Showing 1 to 5 of 1,000 entries

Previous
1
2
3
4
5
...
200
Next

选择预览的数据页码

5、方法选择与计算

5.1、海明距离法的操作介绍

(4) 选择估计方法:

☒ 海明距离法
☐ 贝叶斯法
☐ NPE法
☐ MNPE法
☐ TwoStage法

选择“海明距离法”

点击“开始估计”进行计算

5.2、贝叶斯方法操作介绍

(4) 选择估计方法:

☐ 海明距离法 ☒ 贝叶斯法 ☐ NPE法 ☐ MNPE法 ☐ TwoStage法

选择“贝叶斯法”

输入MCMC抽样次数:

20000

输入MCMC抽样次数，默认为20000次

输入MCMC内燃值:

10000

输入MCMC内燃值，默认为10000

✓ 开始估计

点击“开始估计”进行计算

5.3、NPE 法操作介绍

(4) 选择估计方法:

☐ 海明距离法 ☐ 贝叶斯法 ☒ NPE法 ☐ MNPE法 ☐ TwoStage法

选择“NPE法”

输入考核的知识技能数目:

5

输入考试的知识技能数目（属性个数）；
如果输入了Q矩阵文件，可以不输入此项

✓ 开始估计

点击“开始计算”进行估计

5.4、MNPE 法操作介绍

(4) 选择估计方法:

☐ 海明距离法 ☐ 贝叶斯法 ☐ NPE法 ☒ MNPE法 ☐ TwoStage法

选择“MNPE”法

输入考核的知识技能数目:

5

输入考试的知识技能数目（属性个数）；
如果已输入Q矩阵文件，可不输入此项

✓ 开始估计

点击“开始估计”进行计算

5.5、TwoStage 法操作介绍

(4) 选择估计方法:

☐ 海明距离法 ☐ 贝叶斯法 ☐ NPE法 ☐ MINPE法 ☒ TwoStage法

选择 "TwoStage法"

输入考核的知识技能数目:

5

输入考核的知识技能数目 (属性个数);
如果以输入Q矩阵文件, 可不输入此项

选择阶段IQ并给函数分界规则:

小于等于0.4为0, 大于等于0.4为1

选择阶段-离散规则

输入MCMC抽样次数:

20000

输入MCMC抽样次数, 默认20000次

输入MCMC内燃值:

10000

输入MCMC内燃值, 默认10000

开始估计

点击 "开始估计" 进行计算

6、结果预览与保存

估计结果

下方即为Q矩阵估计结果

Show 5 entries

选择每页显示的行数

	skill1	skill2	skill3	skill4	skill5
1	1	0	0	0	0
2	0	1	1	0	1
3	1	1	1	1	1
4	1	0	0	1	0
5	1	0	0	0	0

Showing 1 to 5 of 31 entries

Previous

1

2

3

4

5

6

7

Next

结果导出

选择想要查看的页数

保存估计的Q矩阵

点击 "保存估计的Q矩阵" 保存计算结果

（三）、Q 矩阵修正和估计的结果解读

结果类型	结果解读
Q 矩阵修正	1. 本模块可实现 DINA、DINO、rRUM、LLM、ACDM、GDINA、Mixed 等模型的 Q 矩阵修正。 2. 结果显示中，带有“#”的位置表明被修正的位置，“0#”表示原始 Q 矩阵此处为 1，修正后为 0；“1#”表示原始 Q 矩阵此处为 0，修正后为 1。
Q 矩阵估计	1. 本模块可实现 DINA 模型的 Q 矩阵估计，后续更多适用模型敬请期待。 2. Q 矩阵估计仅为测验 Q 矩阵界定提供参考，并不能代替测验 Q 矩阵界定过程。

七、图形输出

（一）、图形输出操作示例

1、选择要输出何种图形，项目反应概率图、属性掌握概率散点图或属性掌握概率条形图。



2、上传作答数据以及原始 Q 矩阵

参数设定

(1) 上传作答数据：

Browse... No file selected

点击Browse键上传作答数据文件

(2) 上传Q矩阵：

Browse... No file selected

点击Browse键上传Q矩阵文件

.. ..

	A	B	C	D	E	F	G	H	I	J	K	L
1		Item1	Item2	Item3	Item4	Item5	Item6	Item7				
2	ID1	1	1	1	0	1	1	1				
3	ID2	1	1	1	1	1	1	1				
4	ID3	1	1	1	1	1	1	0				
5	ID4	1	1	1	1	1	1	1				
6	ID5	1	1	1	1	1	1	1				
7	ID6	1	1	1	1	1	1	1				
8	ID7	1	1	1	1	1	1	1				
9	ID8	0	1	1	1	1	1	0				
10	ID9	1	1	1	1	1	0	1				
11	ID10	1	1	1	1	0	0	1				
12	ID11	1	1	1	0	1	1	1				
13	ID12	1	1	1	1	1	1	1				
14	ID13	1	1	1	1	1	1	1				
15	ID14	0	1	1	0	1	1	1				
16	ID15	0	1	1	1	1	1	0				
17	ID16	1	1	1	1	1	1	1				
18	ID17	1	0	1	0	1	1	1				

作答数据格式示例

	A	B	C	D	E	F	G	H	I	J	K	L
1		A1	A2	A3	A4	A5						
2	Item1	1	0	0	0	0						
3	Item2	1	1	1	1	0						
4	Item3	1	0	0	0	0						
5	Item4	1	1	1	1	1						
6	Item5	0	0	1	0	0						
7	Item6	1	1	1	1	0						
8	Item7	1	1	1	1	0						
9	Item8	1	1	0	0	0						
10	Item9	1	0	1	0	0						
11	Item10	1	0	1	1	1						
12	Item11	1	0	1	0	0						
13	Item12	1	0	1	1	0						

Q矩阵数据格式示例

3、选择模型，开始估计

(4) 选择估计模型：

☒ DINA
 ☐ DINO
 ☐ rRUM
 ☐ LLM
 ☐ ACDM
 ☐ GDM
 ☐ LCDM
 ☐ GDINA
 ☐ Mixed
 选择模型

点击“开始估计”按键开始计算

4、在线查看图形，若第一步选择的是项目反应概率图，则需选择在线查看哪道题的反应概率图；若第一步选择的是属性掌握概率散点图或条形图，则要选择在线查看哪个属性的。

⚙️ 输出图形

输入第几题

1

选择输出哪道题的项目反应概率图

⚙️ 输出图形

选择输出哪个属性的掌握概率散点图或条形图

输入第几个属性

1

5、离线保存图形

⚙️ 结果导出

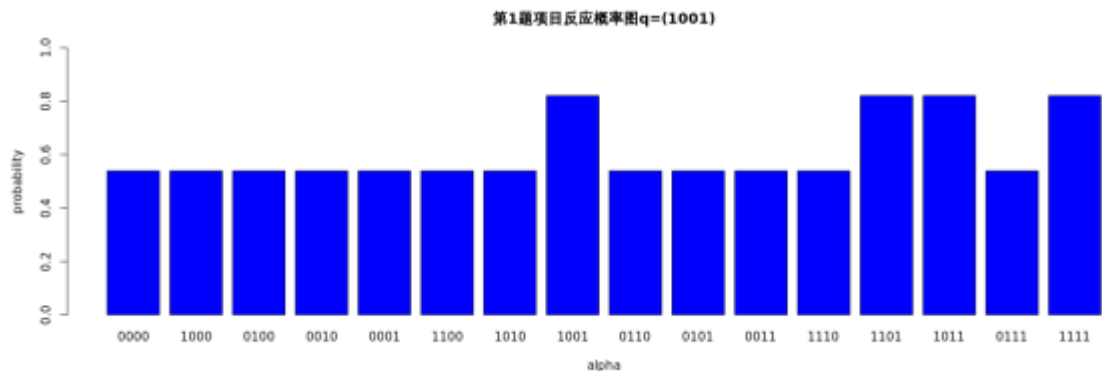
点击按键，离线保存图形

(二)、图形输出结果解读

1、项目反应概率图

输入第几题

1

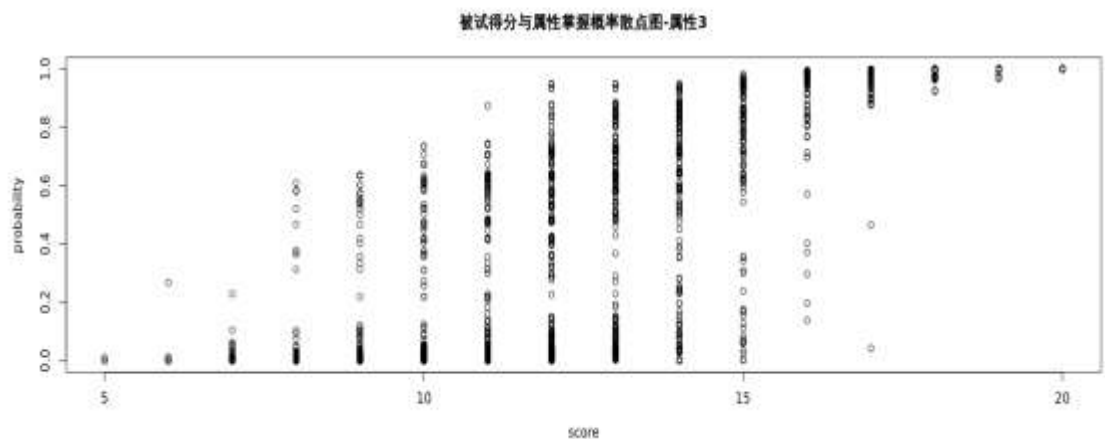


如上面例图所示，项目反应概率图是一种条形图，横坐标为属性掌握模式，纵坐标为答对概率。由例题可知，对第 1 题来说，属性掌握模式为(1001)、(1101)、(1011) 和 (1111) 的被试答对该题的概率最高，大概为 0.8；而属性掌握模式为其它的被试答对该题的概率为 0.6 左右。

2、属性掌握模式散点图

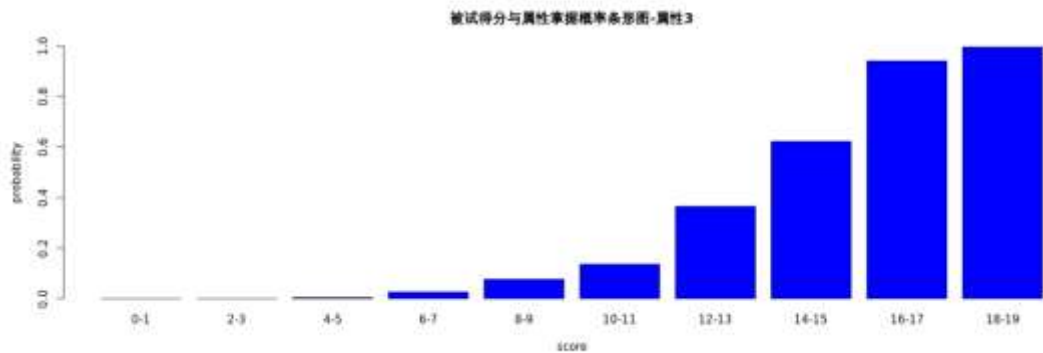
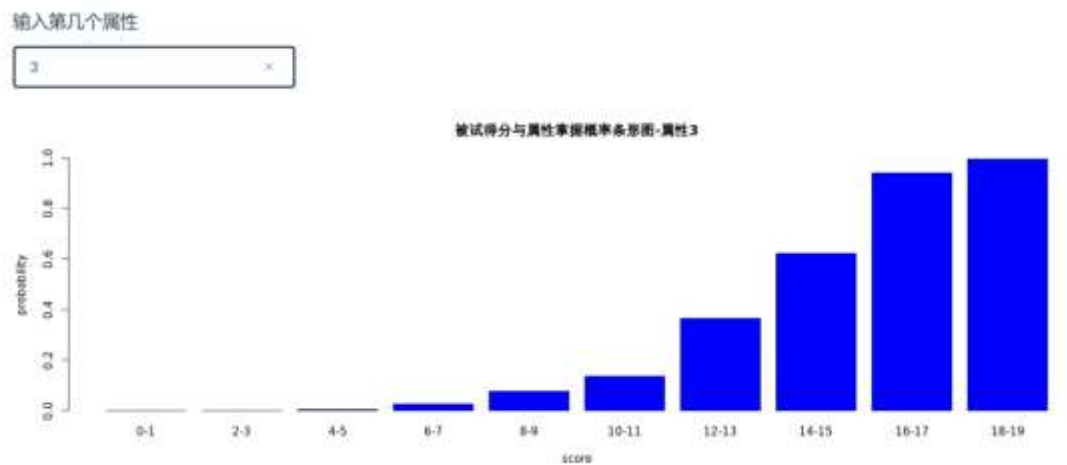
输入第几个属性

3



如上面例图所示，属性掌握模式散点图的横坐标为被试测验总分，纵坐标为掌握属性的概率，属性掌握模式散点图可以更加直观的反映被试总分和属性掌握概率之间的相关关系。由例图可知，对于属性 3 来说，被试测验总分和掌握属性 3 的概率之间总体呈正相关，被试总分越高，则被试掌握属性 3 的概率也越高。

3、属性掌握模式条形图



如上面例图所示，属性掌握模式条形图的横坐标为被试测验总分，纵坐标为掌握属性的概率。由例图可知，随着被试测验总分的增加，被试掌握属性 3 的概率也随之上升。

八、数据服务

(一)、实测数据

1、选择实测数据，本平台提供的示例数据有塔苏卡分数加减法数据、ECPE 英语等级考试数据、TIMSS 八年级测验数据。

(1) 选择示例数据



选择需要下载的数据

2、点击“保存作答数据”和“保存 Q 矩阵”按键来下载相应的作答数据与 Q 矩阵。

⚙ 结果导出

点击“保存作答数据”、“保存Q矩阵”按键来下载作答数据和Q矩阵



(二)、模拟数据

1、点击 Browse 键上传被试属性掌握模式和 Q 矩阵数据，被试属性掌握模式数据非必须，也可以通过系统模拟，上传的数据要求是 CSV 文件。被试属性掌握模式数据第一行为认知属性序号，第一列为被试序号，一行代表一个被试，一列代表一个属性；Q 矩阵文件第一行为认知属性序号，第一列为题目序号，一行代表一个题目，一列代表一个属性。

(1) 上传属性掌握模式：

上传各个被试的属性掌握模式

Browse...

No file selected

(2) 上传Q矩阵：

上传各题目的Q矩阵

Browse...

No file selected

	A	B	C	D	E	F	G
1		A1	A2	A3			
2	ID1	1	1	0			
3	ID2	0	1	0			
4	ID3	1	0	1			
5	ID4	0	0	1			
6	ID5	0	0	1			
7	ID6	0	0	1			
8	ID7	1	0	1			
9	ID8	0	1	0			
10	ID9	0	0	1			
11	ID10	1	0	0			
12	ID11	1	0	1			
13	ID12	1	0	1			
14	ID13	1	0	0			

被试属性掌握模式数据

格式示例

2、选择采用何种认知诊断模型来模拟数据，平台会通过用户选择的模型来计算每个被试在每个题目上的答对概率，再根据答对概率来判定被试在题目上是答对或答错，从而生成相应的作答矩阵。

(4) 选择估计模型：

☒ DINA ☐ DINO ☐ rRUM ☐ LLM ☐ ACDM ☐ GDINA

3、设定模拟的被试人数和属性掌握模式的分布形态（注意，若前面已上传被试

掌握模式数据，则该设定无效)。属性掌握模式的分布形态有两种可供选择，选择“KS 服从均匀分布”，则模拟生成的被试的属性掌握模式是均匀的，即隶属于各个属性掌握模式的被试人数大致相当；若选择“标准正态分布(higher order)”，则平台首先会模拟生成每个被试的能力值（服从标准正态分布），即所有被试的能力均值为 0，被试间的标准差为 1，中等能力的被试人数居多，低能力与高能力被试人数较少，之后再根据能力与 α 的函数（参见 de la Torre et al. [2004] 的 Higher order CDM）模拟产生被试属性掌握模式。

(5) 选择模拟的被试人数：

1000

(6) 选择模拟被试分布：

☒ KS服从均匀分布 ☐ 标准正态分布 (higher order)

4、模拟项目参数，本平台要求使用者对 $P(0)$ 和 $1-P(1)$ 进行设定， $P(0)$ 表示未掌握该项目任何测量属性的被试答对题目的概率， $1-P(1)$ 表示掌握了该项目所有测量属性的被试而答错题目的概率。使用者可选择固定 $P(0)$ 和 $1-P(1)$ ，即使所有题目的 $P(0)$ 和 $1-P(1)$ 为一个固定值；使用者也可使 $P(0)$ 和 $1-P(1)$ 服从均匀分布，通过设定最小值和最大值，在最小值和最大值的区间内为每道题随机抽取数值。

(7) 模拟 $P(0)$ 和 $1-P(1)$ ：

☐ 固定 $P(0)$ 和 $1-P(1)$ ☒ $P(0)$ 和 $1-P(1)$ 服从均匀分布

输入最小值：

0

输入最大值：

0.3

(7) 模拟 $P(0)$ 和 $1-P(1)$ ：

☒ 固定 $P(0)$ 和 $1-P(1)$ ☐ $P(0)$ 和 $1-P(1)$ 服从均匀分布

固定 $P(0)$ ：

0.1

固定 $1-P(1)$ ：

0.1

5、在线查看结果和离线下载结果

⚙️ 模拟结果

📄 模拟作答数据 📄 模拟阅卷系统模式 📄 试题参数 📄 项目反应概率

在线查看结果

⚙️ 结果导出

离线下载结果

📄 保存模拟data

📄 保存7x9010x5

📄 保存试题参数

📄 保存项目反应概率

第三章 参考文献

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Barnes T. (2010). Novel derivation and application of skill matrices: The q-matrix method. *Handbook on educational data mining*, 159-172.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for dichotomously scored items. *Psychometrika*, 35, 179-197.
- Carlin, B. P., & Louis, T. A. (2000). *Bayes and empirical bayes methods for data analysis*. New York, NY: Chapman & Hall
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and Absolute Fit Evaluation in Cognitive Diagnosis Modeling. *Journal of Educational Measurement*, 50, 123-140.
- Cui, Y., & Li, J. (2015). Evaluating person fit for cognitive diagnostic assessment. *Applied Psychological Measurement*, 39(3), 223-238.
- Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement*, 49(1), 19-38.
- de la Torre, J. (2009). DINA Model and Parameter Estimation: A Didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- de la Torre, J., & Lee, Y. S. (2013). Evaluating the wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50, 355-373.
- de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45(4), 346-362
- de la Torre & Chiu. (2016). A General Method of Empirical Q-matrix Validation. *Psychometrika*, 81, 253-273.

-
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- de la Torre, J. (2009). DINA model parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115–130.
- DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, 36(6), 447-468.
- Gierl, M. J., Cui, Y., & Zhou, J. (2009). Reliability and Attribute-Based Scoring in Cognitive Diagnostic Assessment. *Journal of Educational Measurement*, 46(3), 293-313.
- Hartz, S. M. (2002). A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. (Unpublished doctoral dissertation. University of Illinois at Urbana-Champaign.
- Hou, L., de la Torre, J., & Nandakumar, R. (2014). Differential item functioning assessment in cognitive diagnostic modeling: Application of the Wald test to investigate DIF in the DINA model. *Journal of Educational Measurement*, 51, 98-125.
- Holland, P., & Thayer, D. (1986, April). Differential item performance and the Mantel–Haenszel procedure. Paper presented at the American Educational Research Association Annual Meeting, San Francisco, California, USA.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Ma, W., & de la Torre, J. (2016a). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*. 69, 253-275.
- Ma, W., & de la Torre, J. (2016b, July). A Q-matrix validation method for the sequential G-DINA model. Paper presented at the 80th International Meeting of the Psychometric Society, Asheville, NC.
- Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection and attribute classification. *Applied Psychological Measurement*, 40, 200-217.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50-64.

-
- Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement*, 27, 289-298.
- Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (Vol. 7). London: Sage Publications Inc
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: The Guilford Press.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Su, Y.-L., Choi, K. M., Lee, W.-C., Choi, T., & McAninch, M. (2013). *Hierarchical cognitive diagnostic analysis for TIMSS 2003 mathematics*. CASMA Research Report 35. Center for Advanced Studies in Measurement and Assessment (CASMA), University of Iowa.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment*, 327-359.
- Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 51, 337-350.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251-275.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79, 317-339.
- Tutz, G. (1997). Sequential models for ordered responses. In W.J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory*. p. 139-152). New York, NY: Springer.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*. ETS

Research Report RR-05-16. ETS, Princeton, NJ: ETS.

Xiang R. (2013). Nonlinear penalized estimation of true Q-matrix in cognitive diagnostic models[D]. Columbia University.

蔡艳,涂冬波,丁树良. (2013).五大认知诊断模型的诊断正确率比较及其影响因素:基于分布形态、属性数及样本容量的比较. *心理学报*,11:1295-1304.

蔡艳,涂冬波. (2015). 属性多级化的认知诊断模型拓展及其 Q 矩阵设计. *心理学报*,10:1300-1308.

蔡艳,苗莹,涂冬波. (2016).多级评分的认知诊断计算机化适应测验. *心理学报*,10:1338-1346.

蔡艳,谭辉晔,涂冬波. (2015).哪个测验 Q 矩阵更合理:基于 DINA 模型测验 Q 矩阵合理性侦查指标及其比较与应用. *心理科学*,5:1239-1247.

蔡艳,涂冬波,丁树良. (2014). 基于群体水平评估的认知诊断模型开发与应用. *心理科学*, 2:468-472.

丁树良, 祝玉芳, 林海菁, & 蔡艳. (2009). Tatsuoka Q 矩阵理论的修正. *心理学报*, (2),175-181.

刘永. (2015). *基于二阶段测验 Q 矩阵估计方法的开发与应用*.江西师范大学硕士毕业论文.

刘永, 涂冬波. (2015). 认知诊断测验 Q 矩阵估计方法比较. *中国考试*, (5), 53-63.

涂冬波,蔡艳,戴海琦,丁树良.(2010).一种多级评分的认知诊断模型 :P-DINA 模型的开发.*心理学报*,42(10),1011-1020.

涂冬波,蔡艳,戴海琦,丁树良.(2011a). 多维项目反应理论 :参数估计及其在心理测验中的应用. *心理学报*,43 (11) ,1329-1340.

涂冬波,蔡艳,戴海琦.(2012).基于 DINA 模型的 Q 矩阵修正方法. *心理学报*, 44 (4), 558-568.

涂冬波,蔡艳,戴海琦,丁树良. (2012). 一种多策略的认知诊断方法 : MSCD 方法的开发. *心理学报*, 44 (11) ,1547-1533.

涂冬波,蔡艳,戴海琦. (2013).几种常用非补偿型认知诊断模型的比较与选用,基于属性层级关系的考量. *心理学报*,2,243-252.

涂冬波,蔡艳.(2015). 基于属性多级化的认知诊断计算机化自适应测验设计与实现. *心理学报*, 11,1405-1414.

涂冬波, 蔡艳, 戴海琦, 丁树良. (2011). HO-DINA 模型的 MCMC 参数估计及模型性能研究. *心理科学*, 34(6), 1476-1481.

涂冬波, 蔡艳, 戴海琦. (2013). 认知诊断 CAT 选题策略及初始题选取方法. *心理科学*, 36(2), 469-474.

涂冬波,蔡艳,戴海琦,丁树良.(2010b).小学儿童数学问题解决的认知诊断. *心理科学*, 33(6), 1461- 1466.

涂冬波,蔡艳,戴海琦,丁树良.(2011b).HO-DINA 模型的 MCMC 参数估计及模型性能研究. *心理科学*,34(6),1476-1481.

涂冬波,蔡艳,戴海琦,丁树良.(2013). 基于 HO-DINA 模型的多级评分认知诊断模型的开发. *心理科学*, 36(4), 984- 988.

涂冬波,蔡艳,戴海琦,丁树良.(2013),基于项目自动生成的认知诊断测验开发. *心理科学*, 36 (1), 209-214.

涂冬波,蔡艳,戴海琦. (2013). 认知诊断 CAT 选题策略及初始题选取方法. *心理科学*, 36 (2),469-474.

涂冬波,蔡艳,丁树良. (2012).认知诊断理论、方法与应用.北京:北京师范大学出版社.

涂冬波,戴海琦,蔡艳,丁树良. (2010).小学儿童数学问题解决的认知诊断. *心理科学*,33 (6),1461-1466

涂冬波,张心,蔡艳,戴海琦.(2014). 认知诊断资料-模型拟合检验统计量及应用, *心理科学*, 37 (1), 209-214.

汪大勋, 高旭亮, 韩雨婷, 涂冬波. (in press). 一种简单有效的 Q 矩阵估计方法开发 : 基于非参数化方法视角. *心理科学*.