用层级相合性指标探测反应数据中噪音大小*

毛萌萌[1] 丁树良**[2]
（[1]南昌大学公共管理学院心理学系，南昌，330031）（[2]江西师范大学计算机信息工程学院，南昌，330022）

摘 要 不同的认知诊断模型（CDM）对反应数据中噪音的抗干扰能力不同，在评估 CDM 性能的模拟实验中，反应数据中所含噪音的大小是十分重要的实验条件。本文在认知模型已知条件下，欲使用 MHCI 和 NHCI 指标评估认知诊断测验的反应数据（0, 1 评分）中噪音的大小。模拟实验表明，两指标与噪音存在明显的统计规律。尤其是以 NHCI 为主要自变量对噪音进行预测的回归方程中，回归模型解释率均接近 90%；以此实现对噪音的有效预测，从而为选择 CDM 提供一个参考。

关键词 噪音 预测 层级相合性指标 回归方程

1 引言

在认知诊断（cognitive diagnosis, CD）中，认知模型（cognitive model, CM）和认知诊断模型（cognitive diagnostic model, CDM）是两个重要的概念。这两个模型之间有联系，最理想的情形是如同针对不同的测验数据开发不同的项目反应理论模型那样（van den Linden & Hambleton，1997），在认知模型正确条件下，针对不同的认知模型开发相应的 CDM。认知诊断评估即是通过 CDM 从反应数据中挖掘被试不可观察的属性掌握情况，进行补救，以促进其发展。每开发一个 CDM，就要讨论模型的性能表现，甚至要和现存的同类型的模型进行比较、筛选，而这就应该做模拟实验和实证研究。

如果被试反应是理想反应，那么被试对所测的属性掌握情况比较容易判断。比如在一定条件下，精心进行认知诊断测验设计，使得知识状态与理想反应模式一一对应，则从理想反应模式就可以进行判断（丁树良，汪文义，杨淑群，2011；丁树良，杨淑群，汪文义，2010）。然而由于种种原因，比如疏忽大意、疲劳效应、考试压力、猜测等，使得被试反应一定程度上都带有随机性，这种随机性称之为噪音（noise）。不同的分类方法或者 CDM 对噪音抗干扰的能力不同，即有一些诊断模型（方法）在比较小的噪音条件下表现良好，噪音比较大时表现很差，即对噪音干扰不稳健；而有的模型（方法）在比较大的噪音条件下，仍然表现比较好，即对噪音干扰比较稳健（robust）。在其他条件相同时，对噪音稳健的模型适用范围更广。评估诊断模型或方法性能的 Monte Carlo 模拟中，反应数据中噪音的大小是十分重要的实验条件。一般来说，其他条件相同条件下，噪音越小估计准确性越高。在 CDM 开发中，不同研究者在不同的噪音大小条件下进行研究（孙佳楠 , 张淑梅 , 辛涛 , 包钰 , 2011; Cui, Leighton, & Zheng, 2006; de la Torre, 2009; Sun, Xin, Zhang, & de la Torre, 2013）。

问题是选用相应的 CDM 时，因无法直接测量噪音的具体数值，面对实测数据的 CDM 的效率也就不得而知。由于噪音隐藏在作答反应之中而难以分离，所以进行关于噪音大小（或者分布）的假设检验很困难。对确定的属性及层级关系（即认知模型）本文欲使用修改的层级相合性指标 MHCI（丁树良 , 毛萌萌 , 汪文义 , 罗芬 , Cui, 2012）和新的 HCI（NHCI, 毛萌萌 , 2011）评估认知诊断测验的 0-1 反应数据中噪音的大小。MHCI 和 NHCI 都是在层级相合性指标 HCI（hierarchical consistency index, HCI, Cui & Leighton, 2009）基础上修正的。本研究假设是，在其他条件相同的情况下，噪音越小则 MHCI/NHCI 越大；否则，MHCI/NHCI 越小。而这只是一种定性的描述，若能导出并且描述它们之间的统计

* 本研究得到教育部人文社会科学研究青年基金项目“认知诊断中拟合指标及其和噪音关系的研究”（16YJC190016）、国家自然科学基金项目（31360237, 31500909, 31160203）和江西省社会科学规划项目（13JY28）的资助。

** 通讯作者：丁树良。E-mail:ding06026@163.com

DOI:10.16719/j.cnki.1671-6981.20190127

规律性，并且能够导出噪音是影响 MHCI/NHCI 的最重要的因素，那么就可以进一步讨论是否可能由 MHCI/NHCI 来预测反应数据中噪音的大小；若可以，就能够按照噪音的大小寻找最合适使用的 CDM。

2 个人层级相合性指标简介

Cui 和 Leighton（2009）给出第 i 个被试的属性层级指标（HCI_i）为：

$$HCI_i = 1\text{-}2\sum_{j\in SCi}\sum_{g\in S_j} x_{ij}(1-x_{ig})/N_{ci}$$

式中 x_{ij}，x_{ig} 为被试 i 在项目 j，g 上的得分；SCi 表示被试 i 正确反应的项目集合；S_j={g|g 为项目，$AS_g \subseteq AS_j$，$g \neq j$}，称项目 g 为项目 j 的子项目，j 为 g 的父项目，S_j 即项目 j 的子项目集合（j 除外）；N_{ci} 是被试 i 正确作答的项目的子项目数总和。其原理是：设采用 0~1 评分，在不失误的条件下，如果被试 i 能对项目 j 正确反应，则必可对其子项目正确反应；若失误则出现观察反应模式与理想反应模式不一致，用 HCI 度量这种不一致性。

但是 HCI 原始定义中可能出现分母为零的情况，将 HCI 定义中的 S_j 修改为包含项目 j 以纠正这个缺陷，记为 MHCI；注意到 HCI 和 MHCI 只是考虑被试失误，而没有考虑其猜测，为较全面揭示被试反应对理想反应或属性层级的偏离，有 NHCI 指标的提出（丁树良等，2012; 毛萌萌，2011）。

$$NHCI_i = 1-(\sum_{j\in SCi}\sum_{g\in S_j} x_{ij}(1-x_{ig})/N_{ci} + \sum_{k\in SWi}\sum_{f\in S_k} x_{if}(1-x_{ik})/N_{wi})$$

式中 SW_i 表示被试 i 错误反应的项目集合；S_k 即项目 k 的父项目的集合； 是被试 i 错误作答的项目的父项目数总和，其余符号同上。这种既考虑失误又考虑猜测的思想也体现在项目拟合指标（ICI）中（Lai, Cui, & Gierl, 2012）。

在认知模型与 Q 矩阵标定正确的情况下这种差异是由噪音造成的，噪音越大则差异越大，进而层级相合性指标值也越小。而其值越大代表数据与认知模型拟合的越好，因而被广泛用于选择、评价和验证属性层级模型（康春花，辛涛，田伟，2013; 齐冰 ，2008; Gierl, Wang, & Zhou, 2008; Wang & Gierl, 2011）；其中 Wang 和 Gierl（2007）以全部被试的 HCI 均值划分认知模型的拟合程度。

3 实验设计

3.1 从描述到预测之间的逻辑

第 1 节的研究假设中已经定性地描述了噪音大小和 MHCI/NHCI 的关系。然而这种定性的描述在实际应用中有一定困难。必须建立起更加精细的定量关系，即通过 MHCI/NHCI 指标值来预测噪音的值。因此本研究实验主要分为两大部分，实验 1 是通过模拟研究寻找影响 MHCI/NHCI 指标大小的因素（噪音大小是其中重要的因素）；并通过多元线性回归建立起以噪音大小为重要自变量之一，MHCI/NHCI 指标大小为因变量的多元回归方程。实验 2 则是基于逆回归思想（又称逆预测）（Draper & Smith, 1981），在实验 1 明确因果关系的基础上，建立起以噪音大小为待预测变量，MHCI/NHCI 指标大小和实验 1 中的其它影响因素为自变量的逆回归方程，最终实现对不可直接观测的噪音大小的预测。

为了表述清楚，先对文中使用的几个概念进行说明：

（1）理想反应是既不失误也不猜测的作答反应。slippage 指的是被试作答时由于失误、猜测或者其他随机误差引发的与理想反应不同的可能性的大小，slippage 比率即是这种作答反应与理想反应失拟概率。

（2）测验的长度：若 K 表示属性数，Q_r 为缩减 Q 阵（Tatsuoka, 1995, 2009），设 Q_r 是 K 行 m 列 Q 矩阵，测验 Q 矩阵 Q_t 是若干个（比如 L 个）Q_r 堆积而成，即 $Qt=(Q_r,...,Q_r)$。参与构成测验 Q 矩阵的 Q_r 矩阵的个数 L 对测验长度有本质的影响，因为 Q_r 中包含可达阵，可达阵在认知诊断测验设计中起着重要作用（丁树良等，2010, 2011），所以 Q_t 以 Q_r 为测验长度的计量单位，不仅仅考虑了测验长度，而且考虑了测验设计，因此下文中以 Q_r 的个数 L 代替测验的长度。

3.2 实验目的

（1）寻找影响 MHCI 和 NHCI 大小的主要因素（slippage 比率等，下文中用 SP 表示 slippage 比率），并得到它们之间的统计规律；

（2）利用可观察的 K，L 以及可计算的 MHCI 和 NHCI 等因素预测不可观测的 SP。

3.3 实验因素

对 4 种不同属性结构（线性型、收敛型、发散型、独立型）考察不同属性数 K（K=5、6、7）× Qr 的个数 L（L=1、2、3、4）× 不同 SP（SP={.30 .25 .20 .15 .10 .05}）的情况下，被试 MHCI 和 NHCI 均值的整体变化规律。

K=5, 6, 7 时，线性型对应的 Qr 包含的项目类型

分别为5、6、7; 收敛型Q_r包含的项目类型数分别为6、7、8；发散型分别为10、15、25；无结构型作为发散型的特殊情况故本文未考察；对独立型K=5时，Q_r矩阵的项目数为31，K=6、7时，因计算量太大而在以下模拟研究中未考察。这里所说的独立型结构是属性之间互相不为先决 (Tatsuoka, 1995, 2009), 它与上述三种属性层级结构有很大不同。

3.4 实验步骤

以5个属性收敛型，2个Q_r（L=2），SP=.30为例:

（1）依据K和属性层级结构产生Q_r，将Q_r添加全零向量产生学生Q矩阵Qs；（2）根据L对Q_r进行复制以产生测验Q矩阵Q_t，并且依据Q_s重复10遍以产生被试属性矩阵，并据此产生所有被试理想反应模式（IRP）矩阵；（3）根据SP对被试IRP产生失拟以最终生成全体被试得分矩阵X（即所有被试的观察反应模式，简记为ORPs），在此例中SP=.30，即意味着每个被试每道题上的理想反应有.30的可能性产生0、1互变；（4）依据X及Q_t计算所有被试的MHCI和NHCI均值，此类情况重复30遍取平均。

其它情况下得分矩阵可相仿模拟。

4 实验结果

具体实验结果请参见附录，由于篇幅所限，只呈现线性型实验结果。

4.1 描述分析

附录中附表1代表线性型模型下MHCI或NHCI的均值在上述实验条件的变化情况。

以线性型结构为例，如附表1所示，K=5、L=1的情况下，随着SP从.30下降到.05，步长为-.05，MHCI均值从.482逐渐增大至.892，NHCI也有同样的变化趋势。MHCI和NHCI的均值整体上接近，但发散型和独立型较线性型和收敛型更发散。

对于4种层级结构，模拟实验的结果都体现出（1）MHCI或NHCI的均值随着SP的减小而增大;（2）随着L或者K的增多，MHCI或NHCI的均值有减小的趋势。这可能是因为题量越大或属性越多，产生失误的可能性更大。

4.2 回归分析

因变量是MHCI（NHCI）均值，自变量是SP、L、K。以附表1为例，SP从.30下降到.05，步长为-.05，L=1、2、3、4和K=5、6、7条件下，进行逐步回归分析结果见附表2。各种结构MHCI（NHCI）均值分别得到三个回归模型，进入回归模型的自变量均依次是SP、L和K（由于计算量的原因，独立型认知模型的模拟实验中K=5，K在这种情况下为常数，因此进入回归模型的自变量只有SP和L），各回归模型方差分析均显著，回归系数均显著。

线性型结果：如附表2所示，对于MHCI均值来说，回归模型1解释率可以达到.657，模型2达到.917；模型3的解释率可以达到.923。而根据Cohen（1969）提出的效应量指标f2，并且使用计算公式（郑昊敏，温忠麟，吴艳，2011）进行分析。模型1中SP的效应大小为1.915，模型2中L的效应大小为3.133，模型3中K的效应大小为.065；根据Cohen（1969）提出的划分效应大小的标准（.02、.15、.35分别对应效应值的小、中、大），SP和L为比较大的效应值，而K则为比较小的效应值。综合来看模型2更优，得到模型2的回归方程为MHCI均值=1.002-1.844*SP-.089*L。

同理NHCI的回归分析结果见附表2，较优的模型2的回归方程为NHCI均值=1.002-1.842*SP-.089*L。

将四种认知模型中较适宜的回归模型筛选（剔除小效应量的自变量）汇总成表1，表中分别显示各认知模型下因变量为MHCI（NHCI）均值的回归方程，以及第一、第二、第三分别进入的自变量SP、L、K的解释率和效应值。关于收敛型、发散型和独立型的具体结果类似，就不再赘述了。

整体来看，在线性型与收敛型结构的回归分析中，MHCI与NHCI的结果非常相似。对于发散型与独立型，MHCI与NHCI的结果已经出现一定程度的不同，如两类回归模型的解释量上也有较大的不同。

4.3 预测分析

通过上面的定性与定量分析，可看出SP和L会显著的影响拟合指标MHCI或NHCI，K有一定影响但是一般情况下效应量不大。SP是反映数据质量的一个普遍通用的指标，可是SP不可直接测量，能够观察或计算的只有L、K和拟合指标，本研究以SP为因变量，而以MHCI、NHCI的均值、L、K为自变量进行逐步回归。各种结构导出三个回归模型，依次进入回归方程的变量是MHCI（NHCI）均值（为避免共线性两者只能取一）、L和K（独立型认知模型的模拟实验中K=5，K为常数因而不进入此回归方程），各回归模型方差分析均显著，回归系数均显著。

表 1 各结构型 MHCI 和 NHCI 均值回归分析筛选表

	较适宜的回归方程	SP 方程贡献率（效应值）	L 方程贡献率（效应值）	K 方程贡献率（效应值）
线性型	MHCI 均值=1.002-1.844*SP-.089*L	.657(1.915)	.260(3.133)	
	NHCI 均值=1.002-1.842*SP-.089*L	.656(1.907)	.261(3.145)	
收敛型	MHCI 均值=.968-1.827*SP-.083*L	.678(2.106)	.241(2.939)	
	NHCI 均值=.970-1.819* SP-.081*L	.689(2.215)	.232(2.937)	
发散型	MHCI 均值=1.189-1.739* SP-.061*L-.067K	.676(2.086)	.143(.790)	.092(1.045)
	NHCI 均值=1.027-1.791*SP-.048*L-.026*K	.832(4.952)	.101(1.507)	.016(.307)
独立型	MHCI 均值=.434-1.188* SP-.063*L	.513(1.053)	.249(1.046)	
	NHCI 均值=.706-1.510*SP-.040*L	.827(4.78)	.102(1.437)	

表 2 各结构模型对 SP 回归分析筛选表

	较适宜的回归方程	MHCI(NHCI)方程贡献率（效应值）	L 方程贡献率（效应值）	K 方程贡献率（效应值）
线性型	SP=.502-.482*MHCI 均值-.043*L	.657(1.915)	.231(2.063)	
收敛型	SP=.497-.494*NHCI 均值-.040*L	.689(2.215)	.209(2.049)	
发散型	SP=.550-.526*NHCI 均值-.025*L-.014*K	.832(4.952)	.093(1.24)	.017(.288)
独立型	SP=.444-.610*NHCI 均值-.025*L	.827(4.78)	.094(1.19)	

线性型结果：由附表 3 可见回归模型 1，2，3 的解释率分别可达 .657， .888， .895。而 MHCI 均值、L 和 K 的效应值分别为 1.915， 2.063， .067；可见 MHCI 均值和 L 效应值较大，而 K 效应值较小。综合来看模型 2 更优，得到模型 2 的回归方程为 SP=.502-.482*MHCI 均值 -.043*L。

剔除小效应量的自变量后，将各认知模型预测分析中的回归模型汇总成表 2，表中分别显示各认知模型下因变量为 SP 的回归方程，以及第一进入的自变量 MHCI 或者 NHCI 均值（两指标通过模型竞争二选一）、第二进入的自变量 L 和第三进入的自变量 K 和它们的解释率和效应值。关于收敛型、发散型和独立型的具体结果就不再赘述了。

4.4 公式应用示例

以下用两个例子演示表 2 中公式如何使用。

（1）以独立型 5 属性层级结构为例，如果测验的试题矩阵为 2 倍的 Qr（即 L=2），所有被试的 NHCI 均值为 .50（这是比较好的数据）（Wang & Gierl, 2007）。根据表 2 中独立型公式 SP=.444-.610*NHCI 均 值 -.025*L 可 得，SP=.444-.610*.50-.025*2 =.089，则对应该批被试的 slippage 比率为 .089。这个失误比例比较小。

（2）以著名的 Tatsuoka (2002) 分数减法测验数据（含 536 个被试，20 个项目）为例，测验 Q 矩阵使用 de la Torre 和 Douglas (2004) 采用的 Q 阵。使用 Qt 行逐对比较的方法（Tatsuoka, 1995, 2009），推出的层级结构图见附图 2，由此推出 Qr 所含项目数为 65 个。

此批数据中 L=1，K=8，所有被试的 NHCI 均值为 .5876；根据表 2 中发散型公式 SP=.550-.526*NHCI 均 值 -.025*L-.014*K 可 得，SP=.550-.526*.5876-.025*1-.014*8=.104。

当然，对于这个实测数据，如果 L 采取实测数据的测验长度 20 计算则，L=20/65=.308， SP=.550-.526*.5876-.025*.308-.014*8=.121。

这里，请注意两点：第一，本研究根据推导的层级关系图得出 Qr 含 65 项目，未必符合依据属性定义而导出的关系；第二，为佐证本文的方法是否能用，使用 DINA 模型分析这批数据，得出所有题目的 s 和 g 的均值分别是 .128 和 .105，与本研究的预报值相距不远。

5 总结与讨论

（1）本文按照隐式认知诊断模型的代表属性层级方法（AHM，Leighton, Gierl, & Hunka, 2004）的方式模拟被试作答反应，得出预报方程（SP 对 MHCI/NHCI，L，K 的回归方程）；另外按照显式认知诊断模型 DINA 模拟产生作答反应，然后用预报方程对 SP 进行预报，发现预报效果很好，因此认为本文的结果对于非补偿 0-1 评分的认知诊断模型具有一定的推广性。

（2）若获得实测数据并已知 Qt，就可计算 MHCI 和 NHCI（下文将 HCI，MHCI，NHCI 统称为 HCI 类指标）。计算 HCI 类指标后，代入表 2 中回归方程，本文的研究结果显示，就可大致了解噪音的大小，为实际工作者选择认知诊断模型提供帮助；必须注意的是，HCI 类指标广泛适用于各种层

级结构。第 4 节的例子似乎表示，不使用本文的方法而直接使用 DINA 模型也可以获得噪音的大小，但是本研究的初衷是不盲目试用 CDM 的前提下，为挑选合适的 CDM 提供一部分依据（除噪音大小之外，还有模型资料拟合检验、运算速度、数据量大小，等等，均可影响模型选择）。

（3）slippage 比率（SP）和 Qt 中使用 Qr 矩阵的个数 L 对 HCI 类指标的大小有显著的影响，其中 SP 是最主要的影响因素，SP 越大或者 L 越大，HCI 类指标越小；大部分情况下 K 能进入回归方程，但效应值较小。线性型和收敛型下，SP 和 L 能解释 MHCI 或 NHCI 均值变异的 90% 以上；而发散型和独立型情况下这两个指标之间的差距开始增大，此时 NHCI 依旧能有非常好的表现，解释率均有 90% 以上，而 MHCI 则有下滑，尤其是独立型中 SP 和 L 对 MHCI 均值方差的解释率只有 76.2%。可见属性层级越松散（杨淑群 , 蔡声镇 , 丁树良 , 林海菁 , 丁秋林，2008）对认知加工的流畅性越有影响。

（4）以可测因素 MHCI 和 NHCI 均值，以及 L 等做自变量来预测不可观测的 SP，各个认知模型下的回归方程均显著。其中线性型依次进入回归方程的是 MHCI 均值、L 和 K，但 K 的效应较小；因 NHCI 和 MHCI 在线性型模型下表现的高相似性，NHCI 并未进入回归方程（若将 MHCI 替换为 NHCI，回归模型的效果几乎一致）。收敛型依次进入方程的自变量是 NHCI 均值、L 和 K，且 K 的效应较小；此外无论线性型还是收敛型第一个因子的方差解释率都有 65% 以上，两个因子的解释率接近 89%；两者的预测模型的也极为相似。发散型的情况下 NHCI 的优势更明显，进入回归方程的因子依次是 NHCI、L 和 K；而独立型进入回归方程的因子依次是 NHCI、L，其中 K 为常数而未进入回归方程，这两类回归模型第一个因子的解释率均在 83% 左右，两因子的模型解释率有 92% 以上。

综上，如面对实际数据，在实验条件相似的情况下，使用上面两因子的回归模型估计 SP 时就可达到较好的效果。整个实验最少的反应矩阵是 60 人 ×5 题，最大的反应矩阵是 320 人 ×124 题，各个回归方程是在包含 5-7 个属性、1-4 倍 Qr 对应的试题情况下得出的，所以整个结果还是具有一定的稳健性，且无论数据的规模多大，都能快速得到比较准确的预测值。

以往直接使用 HCI 均值描述认知模型和数据的拟合程度不太合适，本文认为影响 HCI 类指标均值的因素还有很多（如噪音，测验长度和属性个数），若将 MHCI（NHCI）通过回归方法得到具体的失拟水平（SP）后，以 SP 指标衡量认知模型和数据的拟合程度似更合适。

因 CDM 的判准率随失拟程度升高而下降，故我们推论那些不能正确判断的被试相对失拟程度较高。面对实测数据并结合 HCI 类指标以推测 SP，进而得到在这种失拟水平下 CDM 的判准率。根据判准率给出划界范围（如模式判准率是 70%，则个人拟合指标最低的 30% 的被试可能误判），有针对性地关注于那些可能被误判的被试，可能减少误判。

国内研究者拓展多级计分的 HCI 指标（康春花 , 吴会云 , 孙小坚 , 曾平飞 , 2018），这为多级计分下预测噪音大小提供了可能。

本文存在一些不足：如被试的失拟按照一定比率随机产生的，但实测数据中有部分被试的失拟率不同且失拟的特点也不一样，此情况下 SP 的估计效果如何有待进一步的研究；试验中的认知模型只有四类，而实际中认知模型多是这四类模型的复合，这时用本文给出的方法是否可用，还是必须重新导出相应的回归方程，这些也有待验证。特别地，在实际应用中，属性层级关系不一定已知，或者不一定正确，Q 矩阵的标注也不一定准确，这时候计算 HCI 类指标的前提缺乏或者不正确，要使用本文介绍的方法就必须想方设法寻找出正确的属性层级关系。HCI 类指标有一定的局限性，其统计分布尚不清楚，所以给出区间估计没有坚实的理论基础，因此本研究没有采用区间估计进行预测，这也是本文的不足之处。本文是在测验 Q 矩阵的元素完全准确条件下做出的结论，这一点在实际应用中要特别注意。

参考文献

丁树良 , 毛萌萌 , 汪文义 , 罗芬 , Cui. (2012). 教育认知诊断测验与认知模型一致性的评估 . *心理学报 , 44(11),* 1535–1546.

丁树良 , 汪文义 , 杨淑群 . (2011). 认知诊断测验蓝图的设计 . *心理科学 , 34(2),* 258–265.

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.

康春花 , 吴会云 , 孙小坚 , 曾平飞 . (2018). 层级一致性指标的多级评分拓展 . *心理科学 , 41(1),* 211–218.

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,* 25–43.

毛萌萌 . (2011). *引进粒计算与形式概念分析技术的认知诊断研究* . 江西师范大学博士学位论文 .

齐冰 . (2008). *HCI 对认知属性层次结构构建失误的侦查研究* . 江西师范大

学硕士学位论文 .

孙佳楠 , 张淑梅 , 辛涛 , 包钰 . (2011). 基于 Q 矩阵和广义距离的认知诊断方法 . *心理学报 , 43(9),* 1095–1102.

杨淑群 , 蔡声镇 , 丁树良 , 林海菁 , 丁秋林 . (2008). 求解简化 q 矩阵的扩张算法 . *兰州大学学报 (自科版), 44(3),* 87–91.

郑昊敏 , 温忠麟 , 吴艳 . (2011). 心理学常用效应量的选用与分析 . *心理科学进展 , 19(12),* 1868–1878.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cui, Y., & Leighton, J. P. (2009). The hierarchy consistency index: Evaluating person fit for cognitive diagnostic assessment. *Journal of Educational Measurement, 46(4)*, 429–449.

Cui, Y., Leighton, J. P., & Zheng, Y. G. (2006). *S*imulation studies for evaluating the performance of the two classification methods in the AHM. *Paper presented at the annual meeting of the NCME, San Francisco, CA, USA.*

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34(1),* 115–130.

de la Torre, J., & Douglas, J. A. (2004). Higher–order latent trait models for cognitive diagnosis. *Psychometrika, 69(3)*, 333–353.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis.* New York: John Wiley & Sons.

Lai, H., Cui, Y., & Gierl, M. J. (2012). Item consistency index: An item–fit index for cognitive diagnostic assessment. *Paper presented at the annual meeting of the NCME,* Vancouver, British Columbia, Canada.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on tatsuoka's rule–space approach. *Journal of Educational Measurement, 41(3),* 205–237.

Sun, J. N., Xin, T., Zhang, S. M., & de la Torre, J. (2013). A polytomous extension of the generalized distance discriminating method. *Applied Psychological Measurement, 37(7),* 503–521.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 51(3)*, 337–350.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–359). Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (2009). Cognitive assessment: *An introduction to the rule space method*. New York: Routledge.

van den Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer.

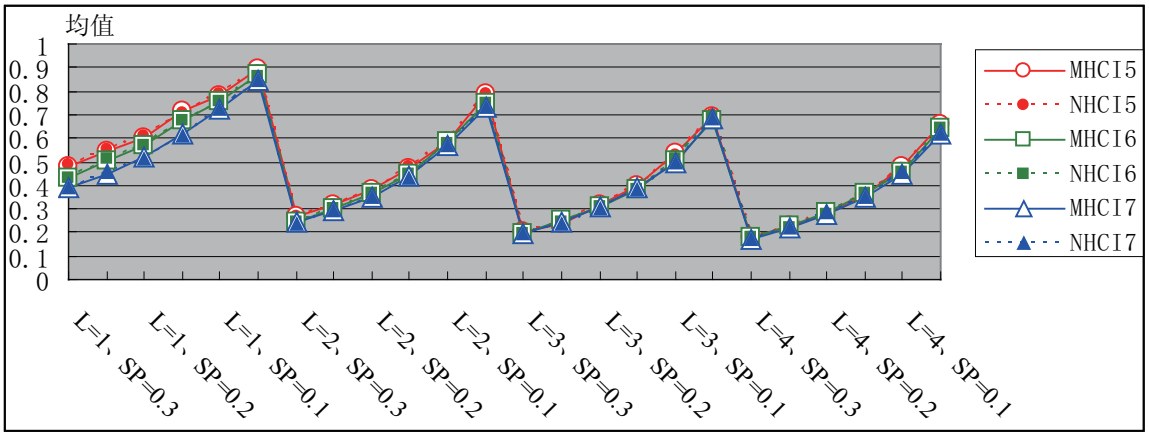
Wang, C. J., & Gierl, M. J. (2007). Investigating the cognitive attributes underlying student performance on the SAT® critical reading subtest: An application of the Attribute Hierarchy Method. *Paper presented at the annual meeting of the NCME*, Chicago, USA.

Wang, C. J., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement, 48(2),* 165–187.

附录 1 线性型实验结果

附表 1 线性型 MHCI 和 NHCI 均值统计

		5 属性		6 属性		7 属性	
L	SP	MHCI	NHCI	MHCI	NHCI	MHCI	NHCI
1	.30	.482	.488	.429	.431	.389	.393
	.25	.549	.551	.506	.510	.447	.453
	.20	.605	.606	.565	.567	.517	.513
	.15	.713	.709	.674	.674	.618	.613
	.10	.784	.785	.755	.753	.724	.727
	.05	.892	.889	.869	.865	.847	.848
2	.30	.265	.265	.246	.248	.243	.232
	.25	.314	.313	.303	.303	.289	.290
	.20	.385	.385	.369	.367	.349	.353
	.15	.473	.479	.448	.450	.436	.434
	.10	.585	.589	.583	.580	.566	.571
	.05	.791	.790	.749	.748	.733	.731
3	.30	.202	.201	.195	.195	.192	.195
	.25	.237	.247	.249	.248	.244	.239
	.20	.321	.317	.312	.308	.306	.305
	.15	.396	.394	.381	.383	.387	.386
	.10	.538	.525	.505	.511	.498	.492
	.05	.693	.703	.674	.682	.679	.680
4	.30	.176	.178	.183	.181	.171	.174
	.25	.228	.225	.226	.221	.223	.223
	.20	.279	.283	.283	.276	.279	.281
	.15	.358	.364	.365	.366	.350	.348
	.10	.478	.470	.452	.458	.450	.454
	.05	.656	.653	.645	.644	.622	.620



附图 1　线性型 MHCI 和 NHCI 均值统计图

如附图 1 所示，标“○”符号的线代表 5 个属性 MHCI 均值的变化趋势，标“□”符号的线代表 6 个属性，标“△”符号的线代表 7 个属性。用 MHCI5 表示含有 5 个属性的 MHCI 均值，MHCI6 表示含有 6 个属性的 MHCI 均值，MHCI7 表示 7 个属性的 MHCI 均值。其他符号类似。

附表 2　线性型 MHCI 和 NHCI 均值回归分析表

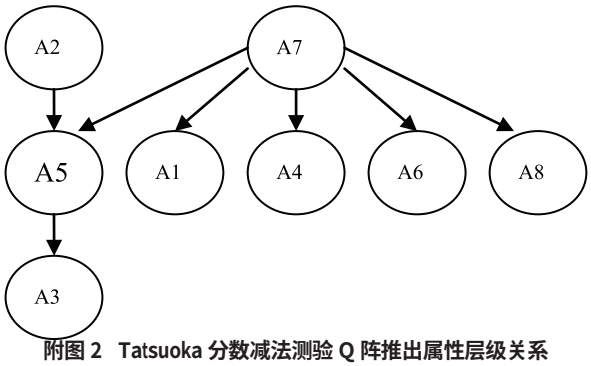
因变量	模型编号	R^2	调整 R^2	$\triangle R^2$	*Sig.F*更改	f^2	预测变量	非标准化 *B*	系数*sig*	*VIF*
MHCI均值	1	.657	.653	.657	.000	1.915	常量	.780	.000	1.000
							SP	-1.844	.000	
	2	.917	.915	.260	.000	3.133	常量	1.002	.000	1.000
							SP	-1.844	.000	1.000
							L	-.089	.000	
	3	.923	.919	.005	.033	.065	常量	1.106	.000	
							SP	-1.844	.000	1.000
							L	-.089	.000	1.000
							K	-.017	.033	1.000
NHCI均值	1	.656	.651	.656	.000	1.907	常量	.780	.000	1.000
							SP	-1.842	.000	
	2	.917	.914	.261	.000	3.145	常量	1.002	.000	1.000
							SP	-1.842	.000	1.000
							L	-.089	.000	
	3	.922	.919	.006	.030	.077	常量	1.109	.000	
							SP	-1.842	.000	1.000
							L	-.089	.000	1.000
							K	-.018	.030	1.000

附表 3　线性型 SP 回归分析表

因变量	模型编号	R^2	调整 R^2	$\triangle R^2$	*Sig.F*更改	f^2	预测变量	非标准化 *B*	系数*sig*	*VIF*
SP	1	.657	.653	.657	.000	1.915	常量	.338	.000	1.000
							MHCI均值	-.356	.000	
	2	.888	.885	.231	.000	2.063	常量	.502	.000	1.351
							MHCI均值	-.482	.000	1.351
							L	-.043	.000	
	3	.895	.890	.007	.044	.067	常量	.555	.000	
							MHCI均值	-.485	.000	1.361
							L	-.043	.000	1.354
							K	-.008	.044	1.007

由于篇幅所限，附录中只呈现线性型结果，如需要完整数据结果，可给 jie_fang1@aliyun.com 邮箱发邮件。

附录 2 Tatsuoka 分数减法测验 Q 阵导出属性层级



附图 2 Tatsuoka 分数减法测验 Q 阵推出属性层级关系

Using Hierarchical Consistency Indexes to Evaluate the Size of the Noise in the 0-1 Response Data of Cognitive Diagnostic Test

Mao Mengmeng[1], *Ding Shuliang*[2]

([1]Department of psychology, School of Public Administration, Nanchang University, Nanchang, 330031)

([2]School of computer information engineering, Jiangxi Normal University, Nanchang, 330022)

Abstract Assessing the performance of cognitive diagnosis model (CDM) in the simulation experiments, the size of the noise in the response data is a very important experimental condition. Due to the hidden noise, it becomes difficult to choose the corresponding CDM in applications. In this article, the modified hierarchical consistency index (MHCI) and new hierarchical consistency index (NHCI) are used to evaluate the size of the noise in the 0-1 response data of cognitive diagnostic test. In order to predict the size of the noise in the response data, Monte Carlo simulation experiment is carried out to find quantitative regularity between the indexes (MHCI, NHCI) and the noise.

Provide that the reduced Q matrix (Q_r) is with K-row and M-column, and the test Q matrix Q_t is pile of L-matrix Q_r, that is $Q_t=(Q_r,...,Q_r)$, where L=1, 2, 3, 4, respectively. L influences the test length essentially. The experiment investigates the changes of the mean values of MHCI (MVM) or NHCI (MVN)) regulated with different attribute structures (linear, convergent, divergent, unstructured model) under the condition of different numbers of attributes (5, 6, 7) × different numbers of Q_r (L=1, 2, 3, 4) × different sizes of the noise (slippage belongs in {.30 .25 .20 .15 .10 .05}). For unstructured model, simulation experiment is carried out with only 5 attributes. For the other K=6, 7, the amount of computing is too heavy to implement. This means the number of attributes in unstructured model is constant.

First, in order to get quantitative regularity between the indexes(MHCI, NHCI)and the noise, through stepwise regression to build the regression equations with MVM or MVN as the dependent variable, slippage ratio, the number of Qr and the number of attributes as the independent variable. Experimental results show that the slip ratio and the number of Qr significantly affect the MVM or MVN. Slippage ratio is the main influence factor; The greater the slippage ratio or the larger times of tests, the smaller the MVM or MVN will be. Number of attributes in most cases can enter the regression equations, but values of the effect are generally small. Regardless of the kind of attribute structures, the two factor regression models have good explanation rate, especially for NHCI index whose rate is above 90%. There is a stable and significant quantitative regularity between slip ratio and index, which provides a way to predict slippage ratio.

In order to predict the size of the noise hidden in the data, inverse regression with slippage ratio is used as the dependent variable, one of MVM or MVN, the number of Q_r and the number of attributes are used as the independent variable, the regression equations are built through stepwise regression. The experimental results show that the linear model is similar to the convergent model, MVM or MVN, the number of Q_r in the regression models have bigger effect. In these two models, the explanation rate of the first factor variance is more than 65%, the explanation rate of the two factors is close to 89%. The results for divergence and unstructured are similar: MVN and the number of Q_r in the regression models have bigger effect. In these two models, the explanation rate of the first factor variance is close to 83%, the explanation rate of the two factors is more than 92%. In conclusion, in experimental conditions of similar cases, just using the above two factor regression models to estimate the slippage ratio can achieve good effect.

Key words size of the noise, prediction, hierarchical consistency index, regression equation