

文章编号: 1000-5862(2015)05-0453-06

GDD-P 在进位计数制中的应用

祝玉芳

(上饶师范学院小教分院 江西 上饶 334000)

摘要:以“进位计数制”为测验内容,运用基于多级评分的广义距离法(简记为GDD-P)进行认知诊断,得到被试的属性掌握模式,这不仅验证了可以对多级评分测验进行认知诊断,且提供了一种多级评分认知诊断的应用范例。

关键词:多级评分;认知诊断;进位计数制

中图分类号: B 842.1 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2015.05.03

0 引言

0.1 问题提出

随着社会的进步,教育工作者在教学过程中,不仅仅只看教育结果,也重视教育过程。教育测验除了作为选拔的功能以外,还可以作为辅助教学和认知诊断的工具。所以,测验除了给一个简单的总分外,还应该给教师以及被试提供被试对知识点的掌握情况,这样,教师的补救教学可以有针对性地进行,被试也可以有目的地学习,从而提高效率,同时教师也可以从测验诊断结果反思教学,从而改进教学。

认知诊断(Cognitive Diagnosis, CD)是根据被试对测试题目的作答情况,通过认知诊断模型计算出被试对测验试题中所涉及各个认知属性的掌握情况,据此给被试提供诊断信息,即通过测验来诊断被试的知识状态。当然,大规模诊断必须借助计算机技术,认知诊断是计算机辅助教学中最重要的内容之一,也是智能教学系统的重要组成部分。

教育测量专家们提出了多种认知诊断模型,这些认知诊断模型通过被试在试题上的作答情况推测被试的知识状态或心理特质。其中开发较早且有一定影响的认知诊断模型是 K. K. Tatsuoka^[1]的规则空间模型(Rule Space Model, RSM),后来 B. M. Jun-

ker等^[2]提出了DINA模型,S. Hartz等^[3]提出了融合模型(Fusion Model),J. P. Leighton等^[4]提出了属性层次模型(Attribute Hierarchy Model, AHM)等。

目前对于认知诊断的应用研究大多建立在2级评分(即0-1评分)模式下,但在测验题型中不仅仅使用2级评分,而且还使用了多级评分,如计算题、论述题等。在许多测验中,0-1评分和多级评分经常混合在一起使用。如果把多级评分模型转化为0-1评分模型来诊断会缺失信息,而0-1评分却可以作为多级评分的一种特例。和2级评分相比,多级评分认知诊断的相关研究较少,特别在应用方面成果就更少。因此,开发一个面向应用的多级评分模型的认知诊断系统是很有必要的。

本文采用多级评分的广义距离法(A polytomous extension of the generalized distance discriminating method, GDD-P)^[5]诊断被试在计算机基础中的“进位计数制”这个内容的掌握情况,开发一个面向应用的多级评分模型的认知诊断系统。

0.2 多级评分的广义距离判别法(GDD-P)

多级评分的广义距离法(GDD-P)是一种多级评分的认知诊断模型的方法。多级评分的广义距离法通过计算被试的观察反应模式和每个理想作答反应模式之间的广义距离,然后将广义距离最小的理想反应模式对应的知识状态作为该被试的知识状

收稿日期: 2015-03-19

基金项目: 国家自然科学基金(31500909, 30860084, 31160203, 31100756, 31360237), 教育部人文社会科学研究青年基金(13YJC880060), 江西省教育科学“十二五”规划2015年度重点课题(15ZD3L045)和江西省教育厅科技计划(GJJ13207, GJJ13208, GJJ13209, GJJ13226, GJJ13227)资助项目。

作者简介: 祝玉芳(1982-),女,江西广丰人,讲师,主要从事计算机辅助教学及教育和心理测量方面的研究。

态.在多级评分的认知诊断测验中,被试*i*和第*t*种理想反应模式的广义距离的数学表达式为

$$d(Y_i, I_t) = \sum_{j=1}^J d(Y_{ij}, I_j^{(t)}), \quad (1)$$

其中

$$d(Y_{ij}, I_j^{(t)}) = |Y_{ij} - I_j^{(t)}|(1 - P_{fj}(\theta_i)), \quad (2)$$

$d(Y_i, I_t)$ 为 Y_i 到 I_t 的广义距离, $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{ij})$ 是被试 i 的观察反应模式, $I_t = (I_1^{(t)}, I_2^{(t)}, \dots, I_j^{(t)})$ 为第 t 种理想反应模式, $d(Y_{ij}, I_j^{(t)})$ 为在项目 j 上被试 i 的观察反应 Y_{ij} 与项目 j 上第 t 种理想反应 $I_j^{(t)}$ 的广义距离, $P_{fj}(\theta_i)$ 是能力为 θ_i 的被试 i 在项目 j 上的作答为 $I_j^{(t)}$ 的概率, $P_{fj}(\theta_i)$ 为一种多级评分 IRT 模型(如 GRM 或 GPCM) (2) 式中的被试 i 的能力参数 θ_i 的值用被试的观察反应模式的估计值 $\hat{\theta}_i$ 来计算.把 $d(Y_i, I_t)$ 值最小的期望反应模式 I_t 对应的属性掌握模式作为被试 i 的属性掌握模式^[5].

1 研究方法

1.1 确定认知诊断评估的目标和群体

本文考察的内容为计算机基础中的“进位计数

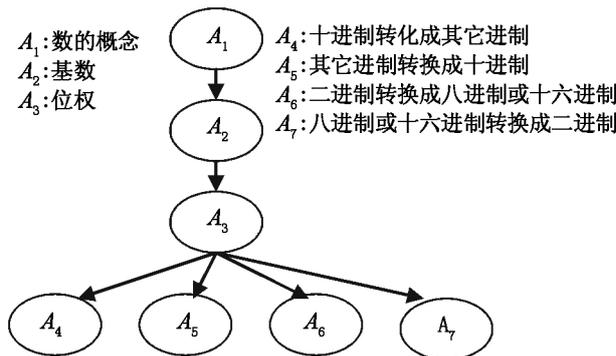


图1 属性层级结构图

制”选取某市某中职学校的学生作为测试对象.测试时间为 45 min,参加测试的学生共 750 人,收回有效数据 705 份,其中一年级(14 级 1 到 7 班、英语班) 338 人,三年级(12 级 16 班) 51 人,四年级(11 级 1、2、4、7、8、9 和 10 班) 316 人.一年级和三年级学生在施测前 2 周学习完,四年级学生在他们一年级学过该内容,在施测的前一节课简单复习了该内容.参加访谈的有 32 人,每个班选取 2 名学生,最终收回有效数据 30 份.

1.2 Q 矩阵

1.2.1 界定属性及其层级结构 陆云娜^[6]和一线的专职教师对“进位计数制”的知识技能进行了界定,确定了 7 种属性及其层级关系,并根据属性层级关系编制了 16 道测验题目来测试这 7 种属性,喻晓锋等^[7]利用贝叶斯网络对属性间层级结构进行修正如图 1.在文献[6]的诊断测验题中,每道题均包含图 1 的 A_1 (数的概念)、 A_2 (基数)和 A_3 (位权)这 3 个属性,所以本文把图 1 的 A_1 、 A_2 和 A_3 这 3 个粒度比较细的属性合并为一个粒度比较粗的属性——进制的概念(图 2 的 A_1),故本文测验题的属性层级结构图如图 2 所示.

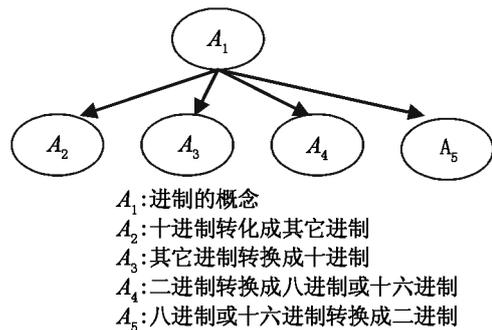


图2 属性层级结构图

1.2.2 Q 矩阵 根据图 2 的属性层级关系可以推导出这 7 种属性的邻接阵,然后从该邻接阵推导出对应的可达阵 R (如表 1),再通过丁树良等^[8-10]的扩张算法或是 K. K. Tatsuoka^[11]的缩减算法导出对应的潜在 Q 阵(Q_r 阵,如表 2), Q_r 阵再加上 1 个 0 列,构成被试 Q 阵(Q_s 阵)^[12].其中 Q_r 阵的列可作为认知诊断测验的项目类, Q_s 阵的列可为被试知识状态的所有可能类.

表1 可达矩阵 R 阵

| | A_1 | A_2 | A_3 | A_4 | A_5 |
|-------|-------|-------|-------|-------|-------|
| A_1 | 1 | 1 | 1 | 1 | 1 |
| A_2 | 0 | 1 | 0 | 0 | 0 |
| A_3 | 0 | 0 | 1 | 0 | 0 |
| A_4 | 0 | 0 | 0 | 1 | 0 |
| A_5 | 0 | 0 | 0 | 0 | 1 |

表 2 Q_r 阵

| | I_1 | I_2 | I_3 | I_4 | I_5 | I_6 | I_7 | I_8 | I_9 | I_{10} | I_{11} | I_{12} | I_{13} | I_{14} | I_{15} | I_{16} |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|
| A_1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| A_2 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 |
| A_3 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| A_4 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| A_5 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |

1.2.3 认知诊断测验题目 本文的认知诊断测验题目在文献 [6] 的基础上修改得到的(如表 1), 主要有以下 2 点改变: (i) 把选择题改成简答题, 并要写解题步骤, 这样减少被试做题时的猜测; (ii) 把 0-1 评分改成多级评分.

该测验假设每个属性的权值为 1, 如第 1 题涉及 2 个属性, 则第 1 题的满分为 2, 从而测验 Q 阵

表 3 测验 Q 阵(Q_r 阵)

| | I_1 | I_2 | I_3 | I_4 | I_5 | I_6 | I_7 | I_8 | I_9 | I_{10} | I_{11} | I_{12} | I_{13} | I_{14} | I_{15} | I_{16} |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|
| A_1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| A_2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A_3 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| A_4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 |
| A_5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

1.2.4 认知诊断方法的选择 基于等级反应模型的属性层级方法(GRM-AHM) [15] 是一种多级评分诊断模型, 应用对数似然比(LL) 作判别方法, 也具有较高的判准率. 基于此, 针对本文的属性层级结构和诊断测验 Q 阵(Q_r 阵), 采用 Monte Carlo 方法对 GDD-P 和 GRM-AHM 进行比较研究, 选择判准率更高的方法作为本测验的认知诊断方法.

具体做法是通过在 4 种被试作答失误率 $slip$ (分别为 2%、5%、10% 和 15%) 的情况下比较 GDD-P 和 GRM-AHM 的判准率, 共 4 个试验条件, 重复 30 次试验以减少误差. 因为本文测验有 705 个被试, 基于考虑到模拟实验尽可能与实测实验条件相似, 故每个试验条件下取被试 705 人. 每个试验都对 2 种诊断方法(GDD-P 和 GRM-AHM) 进行比较.

模拟被试反应模式全集的方法如下: (i) 模拟观察反应模式. 若要模拟每个模式的每个项目的得分有 10% 的概率发生 $slip$ 的情况, 则采用一个服从开区间(0, 1) 上均匀分布 $U(0, 1)$ 的随机数 r , 如果 $r < 0.1$ 且该项目得分不是 0, 则该项目得分减 1 分, 如果 $r > 0.9$ 且该得分不是满分, 则该项目得分增加 1 分, 否则该项目得分不变, 用该方法模拟产生一个以 10% 概率发生 $slip$ 的观察作答项目的得分. 用相同的方法模拟产生 2%、5% 和 15% 失误($slip$) 概率的观察反应模式. (ii) 模拟产生 705 个被试的反应模

(Q_r 阵, 如表 3) 中的值非 0 即 1. 若每个属性的权重不一样, 则 Q_r 阵中的值为对应属性在该题目的权重. 该 Q_r 阵包含可达阵 R , 所以知识状态和期望反应模式一一对应 [13-44]. 测验的期望反应模式全集为 $Q_s^T Q_r$, 即由 Q_s 的转置乘以 Q_r 得到, 其中期望反应全集对应分量的值是多级评分.

式(即模拟产生被试). 把期望反应模式按总得分从小到大排序, 然后使得具有这些得分的被试人数满足标准正态分布, 把得分相同的期望反应模式的个数平均分配 [16].

该模拟实验的归类结果见表 4, 评价诊断方法的优劣指标是用模式判准率和属性边际判准率 [15]. 从表 4 可以看出, 对于本文的测验 Q 阵来说, 总体而言 GDD-P 比 GRM-AHM 的判准率高; 从表 4 也可知道 GDD-P 的属性边际判准率都在 98% 以上, 故本文假设 GDD-P 能准确诊断被试的知识状态. 所以本文采用 GDD-P 对被试进行认知诊断.

表 4 归类结果

| $slip$ | 方法 | 模式判准率 | 属性边际判准率 |
|--------|---------|-----------|-----------|
| 15% | GRM-AHM | 0.811 521 | 0.938 604 |
| | GDD-P | 0.920 291 | 0.981 919 |
| 10% | GRM-AHM | 0.874 283 | 0.955 964 |
| | GDD-P | 0.966 170 | 0.992 397 |
| 5% | GRM-AHM | 0.965 719 | 0.989 239 |
| | GDD-P | 0.986 790 | 0.996 508 |
| 2% | GRM-AHM | 0.971 454 | 0.992 706 |
| | GDD-P | 0.971 068 | 0.989 226 |

数据处理使用 Matlab 7.0 软件编写程序. 测验的项目参数和能力参数是采用 3 参数等级反应模型 [17] 估计的; 使用多级评分的广义距离法(GDD-P) [5] 计算被试的观察反应模式和期望项目反应模

式的广义距离,符合要求的期望反应模式对应的知识状态作为该被试的知识状态,从而得到各被试的属性掌握情况.

2 结果分析

2.1 被试属性掌握情况

被试的属性掌握情况见表 5,从表 5 看出:对所有的被试来说,在 A_1 、 A_2 和 A_3 这 3 个属性上均有 60% 的被试掌握了这些属性,对于 A_4 和 A_5 均有 40% 左右的被试掌握了这 2 个属性.按班级来看,每个班对 A_1 、 A_2 和 A_3 这 3 个属性掌握的被试多于掌握了 A_4 和 A_5 的被试;一般情况下,每个班对 A_1 、 A_2 和 A_3 这 3 个属性掌握的被试人数差不多,对 A_4 和 A_5 掌握的被试人数也不相上下,特别对于 11 级的 7 个班级,每个班对 A_1 、 A_2 和 A_3 这 3 个属性掌握的被试人数均

相同,掌握了 A_4 的被试略多于掌握了 A_5 的被试;11 级 7 个班,每个属性的平均掌握率都高于其他班级的,特别对于前 3 个属性(A_1 、 A_2 和 A_3) 均有 85% 以上的被试掌握了,这说明 11 级这 7 个班的大部分被试都能掌握这 3 个属性.

从认知属性的角度来看,由表 5 知被试对前 3 个属性的掌握情况不相上下.理想情况下,被试对属性 A_1 的掌握率应该高得更多,因为从属性的层级结构来看,属性 A_1 是别的属性的前提,要掌握别的属性必须先掌握属性 A_1 .导致被试对属性 A_1 的掌握率与别的属性相差不大的原因,可能与每道测验题都包含有属性 A_1 有关系,有些题目不会做是因为别的属性没掌握而导致不会做这些题,所以就会导致被试对属性 A_1 总的掌握率就变低了.对于属性 A_1 、 A_2 和 A_3 来说,属性 A_2 的掌握率相对来说会低点,属性 A_2 是十进制转化成其它进制,该转化方法较难理解.

表 5 属性平均掌握率

| 班级 | A_1 | A_2 | A_3 | A_4 | A_5 |
|----------|---------------|---------------|---------------|---------------|---------------|
| 11(1) 班 | 0.918 367 | 0.918 367 | 0.918 367 | 0.530 612 | 0.448 980 |
| 11(2) 班 | 0.948 718 | 0.948 718 | 0.948 718 | 0.871 795 | 0.717 949 |
| 11(4) 班 | 1 | 1 | 1 | 0.911 111 | 0.844 444 |
| 11(7) 班 | 0.854 167 | 0.854 167 | 0.854 167 | 0.583 333 | 0.562 500 |
| 11(8) 班 | 0.977 778 | 0.977 778 | 0.977 778 | 0.822 222 | 0.777 778 |
| 11(9) 班 | 0.955 556 | 0.955 556 | 0.955 556 | 0.888 889 | 0.800 000 |
| 11(10) 班 | 0.888 889 | 0.888 889 | 0.888 889 | 0.622 222 | 0.488 889 |
| 14(1) 班 | 0.333 333 | 0.333 333 | 0.313 725 | 0.117 647 | 0.117 647 |
| 14(2) 班 | 0.372 549 | 0.372 549 | 0.352 941 | 0.117 647 | 0.078 431 |
| 14(3) 班 | 0.085 106 | 0.042 553 | 0.085 106 | 0.021 277 | 0.021 277 |
| 14(4) 班 | 0.326 087 | 0.282 609 | 0.326 087 | 0.130 435 | 0.086 957 |
| 14(5) 班 | 0.346 939 | 0.346 939 | 0.346 939 | 0.102 041 | 0.061 224 |
| 14(6) 班 | 0.684 211 | 0.684 211 | 0.684 211 | 0.263 158 | 0.157 895 |
| 14(7) 班 | 0.717 391 | 0.630 435 | 0.673 913 | 0.478 261 | 0.521 739 |
| 12(16) 班 | 0.470 588 | 0.450 980 | 0.470 588 | 0.254 902 | 0.235 294 |
| 14 英语班 | 0 | 0 | 0 | 0 | 0 |
| 总平均 | 0.619 858 156 | 0.607 092 199 | 0.614 184 397 | 0.422 695 035 | 0.375 886 525 |

被试在掌握属性个数的情况见表 6.从结果看出,被试掌握属性集中在 0 个、3 个和 5 个上,有 36.88% 的被试掌握了全部的 5 个属性,有 38.01%

的被试一个属性都没掌握;少数被试掌握属性个数在 2 个、3 个和 4 个上的.总体来讲,被试掌握属性的情况不够理想.

表 6 被试掌握属性个数的情况

| 属性个数 | 0 | 1 | 2 | 3 | 4 | 5 |
|-------|-------|---|------|-------|------|-------|
| 被试人数 | 268 | 0 | 9 | 129 | 39 | 260 |
| 百分比/% | 38.01 | 0 | 1.28 | 18.30 | 5.53 | 36.88 |

从被试的角度来看,由表 7 可知,705 名被试主要集中在第 1 和第 17 种期望反应模式上,分别有 268 和 260 名被试属于这两种模式;有 35 名被试的属性掌握模式编号是 12,有 128 名被试的属性掌握

模式编号是 7,其他属性掌握模式的被试人数均较少.

结合表 6 和表 7 可知,除去掌握了全部的 5 个属性和 0 个属性的被试外,掌握了 3 个属性的被试

最多,而这些被试集中在认知属性模式为(11100)上;掌握了4个属性的被试主要集中在认知属性模式为(11110)上.

表7 认知模式判别结果

| 属性掌握模式编号 | 属性掌握模式 | 人数 | 百分比/% |
|----------|--------|-----|-------|
| 1 | 00000 | 268 | 38.01 |
| 2 | 10000 | 0 | 0 |
| 3 | 11000 | 3 | 0.43 |
| 4 | 10100 | 6 | 0.85 |
| 5 | 10010 | 0 | 0 |
| 6 | 10001 | 0 | 0 |
| 7 | 11100 | 128 | 18.16 |
| 8 | 11010 | 0 | 0 |
| 9 | 11001 | 0 | 0 |
| 10 | 10110 | 0 | 0 |
| 11 | 10101 | 0 | 0 |
| 12 | 11110 | 35 | 4.96 |
| 13 | 11101 | 2 | 0.28 |
| 14 | 10011 | 1 | 0.14 |
| 15 | 11011 | 0 | 0 |
| 16 | 10111 | 2 | 0.28 |
| 17 | 11111 | 260 | 36.88 |

2.2 访谈结果

对被试进行访谈的目的是了解被试做题时所采用的解题方法,考察被试在解题过程中是否用到本题包含的属性,当正确作答时,是猜测还是掌握本题包含的属性,当出现错误时是什么原因导致的,是失误还是未掌握本题包含的属性.因此,在访谈时,要启发被试详细描述解题的过程和策略.

通过对被试的访谈结果,发现有些项目是可以利用多种策略的,对于一些未包含在 Q 阵中的被试常用的策略,还有一些被试容易出现的错误,具体见表8.

从表8中可以看出,较多题目涉及到多种解题方法和对属性 A_1 的错误,属性 A_1 为进制的概念,涉及的内容比较广,所以,即使有些题涉及的 A_1 能掌握,但不能保证涉及 A_1 的内容都能掌握,这也是为什么被试对 A_1 的掌握率并不比对别的属性高较多的原因.

表8中关于第10题和第12题中的常见错误是:二进制转化为八进制和十六进制的时候,不足位需添加0,理应在最高位添0,却在最低位添0.这种错误会导致 A_4 的掌握率降低,因为 A_4 是二进制转换成八进制或十六进制,转化方法是3位二进制转化1位八进制,4位二进制转化为1位十六进制,不足位添加0,所以如果犯了这种错误,即使掌握了转

化方法也会做错第10题和第12题.

表8 访谈结果发现的新增答题关键点

| 题号 | 新增答题关键点 |
|----|---|
| 1 | 把所有数都转化为二进制再比较大小,则属性掌握模式为($A_1 A_2 A_3$) |
| 2 | 把十六进制字母为H混淆为其数码 |
| 5 | 十进制转化二进制采用除2取余后,却用顺序排列 |
| 6 | 计算错误 |
| 7 | 把十进制转化为二进制,再把二进制转化为十六进制,则属性掌握模式为($A_1 A_2 A_4$) |
| 8 | 计算错误 |
| 9 | 十进制转化八进制采用除8取余后,却用顺序排列 |
| 10 | 把二进制转化为八进制时,不足三位,应在最高位添0,却在最低位添0(错误) |
| 11 | 把八进制6转化成3位二进制错误 |
| 12 | 把二进制转化为十六进制时,不足四位,应在最高位添0,却在最低位添0(错误) |
| 13 | 误把十六进制D转化为十进制数14 |
| 14 | 把八进制转化为十进制,再把十进制转化为十六进制,则属性掌握模式为($A_1 A_2 A_3$) |
| 15 | 把八进制转化为十进制,再把十进制转化为十六进制,则属性掌握模式为($A_1 A_2 A_3$) |
| 16 | 误把十六进制A转化为十进制数11 |

3 小结与展望

本文运用多级评分的广义距离法对“进位计数制”这一内容进行认知诊断分析,这为如何进行多级评分认知诊断的应用研究提供方法,也为指导“进位计数制”学习、教学及评估提供丰富的信息.

对705名被试的研究表明:

1) 这些被试对“进位计数制”所涉及的5个关键认知属性的掌握的总体情况不够理想,特别是 A_4 和 A_5 的掌握率都不到50%,属性 A_1 、 A_2 、 A_3 的掌握率相对高点,在60%以上.11级的7个班对这前3个认知属性的掌握情况都比较好,均有85%以上,所以教师只需对于属性 A_4 和 A_5 ,以及一些容易出错的地方(表8已列出)再强调巩固即可.尤其值得注意的是14(3)班和14(英)这2个班对这5个认知属性几乎都没掌握,应引起任课教师的重视.

2) 对被试掌握认知属性的个数来看,大部分被试掌握了全部的5个属性和0个属性,然后是掌握了3个属性和4个属性;

3) 根据30个有效被试的访谈结果:看出被试常见的一些错误;有4道测验题可用不同的属性掌

握模式来做也行,即这些题目可用不同的解题策略.

由于本文的有些测验项目是可以多种解题策略的,故对于本文的测验可以用不同的 Q 矩阵来描述,到底是选取其中某一个 Q 矩阵还是采用多个 Q 矩阵可以使得诊断更加准确,值得探讨.对于本文使用的 Q 矩阵如何评估、修正,这是本文并未涉及的一个重要问题,有待进一步探讨和分析.

4 参考文献

- [1] Tatsuo K K. Rule space: an approach for dealing with misconceptions based on item response theory [J]. Journal of Educational Measurement, 1983, 20(4): 345-354.
- [2] Junker B M, Sijtsma K. Cognitive assessment models with few assumptions and connections with nonparametric item response theory [J]. Applied Psychological Measurement, 2001, 25(3): 258-272.
- [3] Hartz S, Roussos L, Stout W. A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality [D]. Urbana: University of Illinois at Urbana-Champaign, 2002.
- [4] Leighton J P, Gierl M J, Hunka S M. The attribute hierarchy method for cognitive assessment: a variation on Tatsuo's rule space approach [J]. Journal of Educational Measurement, 2004, 41(3): 205-237.
- [5] Sun Jia'nan, Xin Tao, Zhang Shumei, et al. A polytomous extension of the generalized distance discriminating method [J]. Applied Psychological Measurement, 2013, 37(7): 503-521.
- [6] 陆云娜. 规则空间模型在进位计数制诊断性测验中的应用 [D]. 南昌: 江西师范大学, 2008.
- [7] 喻晓锋, 丁树良, 秦春影, 等. 贝叶斯网在认知诊断属性层级结构确定中的应用 [J]. 心理学报, 2011, 43(3): 338-346.
- [8] Ding Shuliang, Luo Fen, Cai Yan, et al. Complement to Tatsuo's Q matrix theory [C]. Tokyo: Universal Academy, 2008.
- [9] 丁树良, 祝玉芳, 林海菁, 等. Tatsuo Q 矩阵理论的修正 [J]. 心理学报, 2009, 41(2): 175-181.
- [10] 杨淑群, 蔡声镇, 丁树良, 等. 求解简化 Q 矩阵的扩张算法 [J]. 兰州大学学报: 自然科学版, 2008, 44(3): 87-91.
- [11] Tatsuo K K. Architecture of knowledge structure and cognitive diagnosis: a statistical pattern recognition and classification approach [C]. Hillsdale, NJ: Erlbaum, 1995.
- [12] 丁树良, 汪文义, 罗芬. 认知诊断中 Q 矩阵和 Q 矩阵理论 [J]. 江西师范大学学报: 自然科学版, 2012, 36(5): 441-445.
- [13] 丁树良, 杨淑群, 汪文义. 可达矩阵在认知诊断测验编制中的重要作用 [J]. 江西师范大学学报: 自然科学版, 2010, 34(5): 490-494.
- [14] 丁树良, 罗芬, 汪文义. 认知诊断分类中心的确定 [J]. 心理学探新, 2013, 33(5): 396-401.
- [15] 祝玉芳, 丁树良. 基于等级反应模型的属性层级方法 [J]. 心理学报, 2009, 41(3): 267-275.
- [16] 祝玉芳. RSM 改进及多级评分 AHM 的开发研究 [D]. 南昌: 江西师范大学, 2005.
- [17] 陈青, 丁树良, 朱隆尹, 等. 3 参数等级反应模型及其参数估计 [J]. 江西师范大学学报: 自然科学版, 2010, 34(2): 117-122.

An Application of the GDD-P in Carrying Notation

ZHU Yufang

(Shangrao Normal School, Shangrao Jiangxi 334000, China)

Abstract: Using a polytomous extension of the generalized distance discriminating method (GDD-P) to diagnose students' mastery of carrying notation, the students' attribute mastery patterns are obtained. The study indicated that not only examinees' attribute pattern could be diagnosed with polytomous response, but it also provides an example for cognitive diagnosis with polytomous response.

Key words: polytomous response; cognitive diagnosis; carrying notation

(责任编辑: 冉小晓)