

# 基于多元概化理论的全 国 硕 士 研 究 生 中 医 综 合 考 试 质 量 分 析 \*

白 娟

(教育部考试中心,北京 100084)

关键词:中医综合考试;多元概化理论;试卷结构;测量精度

中图分类号:G424.74

文献标识码:A

文章编号:1673-9043(2014)01-0044-03

全国硕士研究生入学中医综合考试的目的是科学、公平、有效地测试考生是否具备继续攻读中医硕士学位所需要的中医药学基础知识和基本技能。《中医综合考试大纲》规定考试涵盖中医基础理论、中医诊断学、中药学、方剂学、中医内科学和针灸学 6 门课程,设有 A 型、B 型、X 型 3 种题型。对照《考试大纲》,多元概化理论为试卷结构、学科比例分配、信度等问题的研究与论证提供了框架<sup>[1]</sup>。

概化理论(GT)是一种把测量误差作为模型参数来处理的测量理论,是经典测量理论与实验设计、方差分析相结合的产物。根据测量目的的不同,研究者可以设计不同的概括全域,通过多个决策研究计算出不同概括全域上的多个信度指标,尤其是多元概化理论,在处理含有多个潜在能力因子,且因子之间存在相关的测量问题方面有独到之处<sup>[2]</sup>。

本文应用多元概化理论评估 2010—2012 年全国硕士研究生入学中医综合考试的试卷结构和测量精度等,获取“平均”或“趋势”的结果,从整体上评价中医综合考试,并为今后修订《考试大纲》、改进命题方案、提高考试质量提供参考依据。

## 1 资料与方法

**1.1 研究样本** 本研究采取分层等距抽样法,各年度考生抽取人数及考试数据的描述统计见表 1。

**1.2 测量的设计** 根据多元概化理论的研究设计,从测量学科能力的角度,可以将中医综合能力看成是包含中医基础理论、中医诊断学、中药学、方剂

表 1 2010—2012 年中医综合考试的描述统计

| 年份   | 样本量   | 平均分    | 标准差   | 难度    | $\alpha$ 信度 |
|------|-------|--------|-------|-------|-------------|
| 2010 | 3 282 | 157.48 | 38.63 | 0.525 | 0.942       |
| 2011 | 3 050 | 161.17 | 38.49 | 0.537 | 0.939       |
| 2012 | 3 280 | 156.37 | 41.85 | 0.521 | 0.945       |

学、中医内科学、针灸学共 6 个学科因子的单面  $p^0x_i^0$  设计。另外,可以将中医综合考试看成是包含 A 型题、B 型题、X 型题 3 种题型因子的  $p^0x_i^0$  设计。

## 2 研究结果

### 2.1 关于 6 个学科因子的分析

**2.1.1 G 研究** 各效应在 6 种学科因子上的方差协方差分量。

运用 mGENOVA 软件可以得到考生( $p$ )、试题( $i$ )以及考生与试题之间交互效应( $p \times i$ ) 在 6 个学科因子上的方差和协方差分量的估计矩阵,见表 2。

在 6 个学科因子中,2010—2012 年方差分量最大的均是方剂学和中药学,其次为针灸学和中医内科学,中医基础理论和中医诊断学的方差分量相对较低。这说明在研究生入学中医综合考试中,方剂学和中药学发挥作用最大,其次为针灸学和中医内科学,中医基础理论和中医诊断学的作用还应进一步加强。同时,根据协方差分量的估计值,发现连续 3 年各个学科与其它 5 个因子的协方差分量值都小于 0.1,说明考生在各学科所得分数高低的一致性程度相对较低,各学科因子在预测考生能力水平方面发挥各自不同的作用<sup>[3]</sup>。

**2.1.2 D 研究** 各学科因子全域分估计及全域总分估计的精度问题。

按照 6 个学科因子试题量所占比重来决定权重系数,对 6 个学科因子全域分进行合成,可以得到

\* 基金项目:全国教育科学规划 2011 年度教育部重点课题资助项目(GFA111012)。

作者简介:白娟(1980-),女,助理研究员,研究方向为教育考试理论与评价。

表2 G研究中各效应在6个学科因子上的方差协方差分量估计

| 年度        | 效应       | 中基    | 中诊    | 中药    | 方剂    | 中内    | 针灸    |
|-----------|----------|-------|-------|-------|-------|-------|-------|
| 2010      | <i>p</i> | 0.039 | 0.990 | 0.899 | 0.894 | 0.935 | 0.829 |
|           |          | 0.034 | 0.030 | 0.920 | 0.930 | 0.954 | 0.854 |
|           |          | 0.042 | 0.038 | 0.057 | 0.933 | 0.956 | 0.801 |
|           |          | 0.045 | 0.041 | 0.057 | 0.065 | 0.960 | 0.827 |
|           |          | 0.040 | 0.036 | 0.050 | 0.053 | 0.048 | 0.806 |
|           |          | 0.037 | 0.034 | 0.044 | 0.048 | 0.040 | 0.052 |
|           | <i>i</i> | 0.116 | 0.145 | 0.113 | 0.104 | 0.160 | 0.110 |
| <i>pi</i> | 0.483    | 0.440 | 0.506 | 0.499 | 0.434 | 0.479 |       |
| 2011      | <i>p</i> | 0.031 | 0.917 | 0.841 | 0.897 | 0.877 | 0.762 |
|           |          | 0.029 | 0.032 | 0.882 | 0.915 | 0.937 | 0.772 |
|           |          | 0.037 | 0.039 | 0.060 | 0.957 | 0.910 | 0.770 |
|           |          | 0.039 | 0.040 | 0.057 | 0.059 | 0.930 | 0.802 |
|           |          | 0.036 | 0.039 | 0.052 | 0.053 | 0.054 | 0.784 |
|           |          | 0.031 | 0.032 | 0.044 | 0.045 | 0.042 | 0.053 |
|           | <i>i</i> | 0.156 | 0.085 | 0.143 | 0.118 | 0.070 | 0.070 |
| <i>pi</i> | 0.427    | 0.527 | 0.456 | 0.493 | 0.523 | 0.545 |       |
| 2012      | <i>p</i> | 0.040 | 0.971 | 0.847 | 0.861 | 0.917 | 0.805 |
|           |          | 0.040 | 0.043 | 0.897 | 0.913 | 0.971 | 0.825 |
|           |          | 0.044 | 0.049 | 0.068 | 0.932 | 0.888 | 0.759 |
|           |          | 0.052 | 0.057 | 0.074 | 0.092 | 0.939 | 0.792 |
|           |          | 0.038 | 0.042 | 0.049 | 0.060 | 0.044 | 0.824 |
|           |          | 0.040 | 0.043 | 0.050 | 0.061 | 0.044 | 0.064 |
|           | <i>i</i> | 0.093 | 0.078 | 0.105 | 0.067 | 0.002 | 0.058 |
| <i>pi</i> | 0.503    | 0.543 | 0.494 | 0.525 | 0.563 | 0.560 |       |

注:1)关于*p*:主对角线上的元素为各效应在相应子类型上的方差分量估计;主对角线以下元素为各效应在不同因子间协方差分量的估计;主对角线以上元素为因子间的相关系数的估计。2)关于*i*和*pi*:由于假设题目是从各个侧面下随即抽取的,各个维度之间的题目没有相关,是独立的,因此仅报告*i*和*pi*的方差分量,并以横向形式列出。

全域总分的方差以及相应误差的方差分量估计,进而估计全域总分的概化系数,见表3。

表3 考生全域分数等在6个学科因子以及全域总分上估计的方差分量值

| 年度   | 指标     | 中基    | 中诊    | 中药    | 方剂    | 中内    | 针灸    | 全域总分  |
|------|--------|-------|-------|-------|-------|-------|-------|-------|
| 2010 | 全域分数方差 | 0.039 | 0.030 | 0.057 | 0.065 | 0.048 | 0.052 | 0.044 |
|      | 相对误差   | 0.016 | 0.015 | 0.017 | 0.017 | 0.012 | 0.019 | 0.003 |
|      | 概化系数   | 0.706 | 0.671 | 0.773 | 0.796 | 0.793 | 0.731 | 0.943 |
| 2011 | 全域分数方差 | 0.031 | 0.032 | 0.060 | 0.059 | 0.054 | 0.053 | 0.043 |
|      | 相对误差   | 0.014 | 0.018 | 0.015 | 0.016 | 0.015 | 0.022 | 0.003 |
|      | 概化系数   | 0.688 | 0.649 | 0.798 | 0.783 | 0.785 | 0.710 | 0.939 |
| 2012 | 全域分数方差 | 0.040 | 0.043 | 0.068 | 0.092 | 0.044 | 0.064 | 0.051 |
|      | 相对误差   | 0.017 | 0.018 | 0.016 | 0.017 | 0.016 | 0.022 | 0.003 |
|      | 概化系数   | 0.703 | 0.704 | 0.806 | 0.840 | 0.734 | 0.740 | 0.945 |

综合2010—2012年3年的考试数据,方剂学、中药学、中医内科学和针灸学测量的相对误差小,概化系数高,这表明在中医综合考试中,这4门学科测量精度相对较高。中医基础理论和中医诊断学的概化系数介于0.6~0.7之间,其测量精度处于可接受水平,但与其它4门学科相比相对较低,有待进一步提高。总体上,分析3年考后数据,按照6个学科因子合成的全域总分的相对误差很小(小于0.005),概化系数均在0.93以上,表明中医综合考试的整体测量精度比较理想。

## 2.2 关于3种题型因子的分析

### 2.2.1 G研究 各效应在3种题型因子上的方差协方差分量。

运用mGENOVA软件可以得到考生(*p*)、试题(*i*)以及考生与试题之间交互效应(*pxi*)在三种题型因子上的方差和协方差分量的估计矩阵,见表4。

表4 G研究中各效应在3个题型因子上的方差协方差分量估计

| 年度       | 效应        | A型题       | B型题      | X型题   |       |
|----------|-----------|-----------|----------|-------|-------|
| 2010     | <i>p</i>  | 0.035     | 0.982    | 0.989 |       |
|          |           | 0.038     | 0.042    | 0.956 |       |
|          |           | 0.045     | 0.048    | 0.059 |       |
|          | <i>i</i>  | 0.093     | 0.073    | 0.128 |       |
|          |           | <i>pi</i> | 0.419    | 0.346 | 0.693 |
|          |           | 2011      | <i>p</i> | 0.035 | 0.994 |
| 0.040    | 0.047     |           |          | 0.966 |       |
| 0.042    | 0.048     |           |          | 0.053 |       |
| <i>i</i> | 0.079     |           | 0.064    | 0.150 |       |
|          | <i>pi</i> |           | 0.393    | 0.391 | 0.715 |
|          | 2012      |           | <i>p</i> | 0.045 | 0.100 |
| 0.047    |           | 0.048     |          | 0.986 |       |
| 0.052    |           | 0.054     |          | 0.061 |       |
| <i>i</i> |           | 0.085     | 0.077    | 0.103 |       |
|          |           | <i>pi</i> | 0.399    | 0.425 | 0.799 |

由表4可知,2010—2012年连续3年,在3种题型因子中,方差分量最大的是X型题,其次为B型题,最小的是A型题。这说明在中医综合考试中,X型题的作用最大,A型题的作用最小。同时,根据协方差分量的估计值,发现各个题型与其它2个因子的协方差分量值都小于0.1,说明考生在各题型上所得分数高低确定顺序的一致性程度相对较低。

### 2.2.2 D研究 各题型因子全域分估计及全域总分估计的精度问题。

按照3个题型因子试题量所占比重来决定权

系数,对3种题型因子全域分进行合成,可以得到全域总分的方差以及相应误差的方差分量估计,进而估计全域总分的概化系数,见表5。

表5 考生全域分数等在3个题型因子以及全域总分上估计的方差分量值

| 年度   | 指标     | A型题   | B型题   | X型题   | 全域总分  |
|------|--------|-------|-------|-------|-------|
| 2010 | 全域分数方差 | 0.035 | 0.042 | 0.059 | 0.044 |
|      | 相对误差   | 0.005 | 0.009 | 0.012 | 0.003 |
|      | 概化系数   | 0.882 | 0.830 | 0.837 | 0.943 |
| 2011 | 全域分数方差 | 0.035 | 0.047 | 0.053 | 0.043 |
|      | 相对误差   | 0.005 | 0.010 | 0.012 | 0.003 |
|      | 概化系数   | 0.877 | 0.827 | 0.816 | 0.939 |
| 2012 | 全域分数方差 | 0.045 | 0.048 | 0.061 | 0.051 |
|      | 相对误差   | 0.005 | 0.011 | 0.013 | 0.003 |
|      | 概化系数   | 0.901 | 0.820 | 0.822 | 0.944 |

综合2010—2012年3年数据,由于A型题测量误差最小,因而相对于其它两种题型,A型题概化系数最高,测量精度最高。X型题和B型题的相对误差也较小,概化系数在0.8以上,测量精度也较高。总体上,3年中医综合考试各题型全域总分的相对误差都很小,概化系数高,这表明,从题型合成角度来看,整体测量精度非常理想。

### 3 讨论与结论

本文运用多元概化理论分别从学科和题型的角度探讨了全国硕士研究生入学中医综合考试的试卷结构和测量精度问题,结果发现有以下几个方面。

**3.1 从考试内容看中医综合考试** 从考试内容来看,各学科因子的方差和协方差成分的估计是比较稳定的,这说明各个学科试题的命制是严格遵循《考试大纲》的,各个学科的分量分配是基本合适的。相对而言,方剂学、中药学、针灸学和中医内科学的贡献总体较高,中医基础理论和中医诊断学的分量和贡献还需要进一步平衡。从D研究的结果看,中医综合考试的整体测量精度较高,中医基础理论和中医诊断学的测量精度有待进一步加强<sup>[4-6]</sup>。

从概化理论D研究的角度来讲,提高样本容量,信度会有一定程度的提高。但是考虑到中医综合试卷共有180道题,已经是一个比较长的试卷,完全靠增加题量提高信度是不太可行的。对中医基

础理论和中医诊断学的试题本身进行分析发现,这两门学科的试题难度总体上低于整张综合试卷的平均难度,不同水平的考生答对概率相对较高,导致题目的区分度相对较低,为此,适当提高题目难度,提高区分度可能是改进中医基础理论和中医诊断学试题的一个方向。另外,考虑到中医基础理论和中医诊断学的学科本身,有一些知识点存在一定争议,不十分明确的知识点相对其它学科较多,在一定程度上增加了命题难度。

**3.2 从题型设计看中医综合考试** 从题型来看,各题型的方差和协方差成分的估计是比较稳定的,不同年度的试卷在结构上可看成是“平行的”。从D研究的结果来看,按照题型合成分数,由于都是客观选择题型,总体上测量误差小,测量精度非常理想。

中医综合考试连续多年采用的都是客观题的形式,题型相对一致统一,答案精确,评分误差影响很小,因此采用多元概化理论进行试卷分析的适用度很高。但是,由于国家大规模考试的局限性,本研究在使用多元概化理论分析试卷时也存在一定不足,例如,由于是国家统一考试,每年都使用不同的试卷,没有共同题和共同人,无法对不同年度的分数进行等值处理,可能会在一定程度上影响到分数的可比性。

#### 参考文献:

- [1] 张伯礼,于越,周桂彤.中医药高等教育教学改革思路探讨[J].天津中医药大学学报,2008,27(2):87-88.
- [2] 杨志明,张雷.测评的概化理论及其应用[M].北京:教育科学出版社,2003.
- [3] 关丹丹,王博,车宏生.2007—2010年心理学专业基础综合考试的多元概化理论研究[J].心理科学,2011,34(4):950-956.
- [4] 张艳军,于虹,高秀梅.加强实践教学环节,增强药类专业学生创新实践能力[J].天津中医药大学学报,2006,25(3):153-154.
- [5] 于河,王思成,白俊杰,等.中医教育循证决策研究方法与应用举隅[J].天津中医药大学学报,2012,31(3):166-168.
- [6] 张天懿,何强,周桂桐,等.中医学专业认证管理模式研究[J].天津中医药大学学报,2013,32(4):232-233.

(收稿日期:2013-12-10)