

# 小学 1~4 年级学生思维能力测验的初步编制\*

胡卫平 李海燕 秦亚平

(山西师范大学教育科学研究院, 临汾 041000)

**摘 要** 编制了“小学 1~4 年级学生思维能力测验”, 该测验有文字、数字、图形三种测试材料, 包括比较分类、归纳推理、演绎推理、空间认知、类比推理、抽象概括 6 个分测验。793 名 1~4 年级小学生参加了测试, 结果表明该测验难度适中、区分度较好、信度和效度较高。

**关键词** 小学生, 思维能力, 测验编制。

**分类号** G449.5

## 1 问题提出

思维是智力和能力的核心, 是人脑对客观现实的反映<sup>[1]</sup>。研究儿童青少年思维能力的发展与培养, 对于实施素质教育, 培养创造性人才, 具有重要的意义。思维能力的有效评估是思维能力发展与培养研究的前提。目前思维能力测验较少, 虽然国外有一些著名的测量量表, 如《比纳智力量表》和《韦克斯勒智力量表》, 但并非针对思维能力而设置的题目, 而且只适合于个别施测; 《瑞文推理测验》只测查了推理能力。国内也有学者编制了一些思维能力的测验, 如小学儿童的假言推理测验<sup>[2]</sup>, 思维发展水平测验<sup>[3]</sup>, 但是要么材料单一, 测查了思维能力的某一方面, 要么测验时间太长, 不便于操作。基于此, 编制了适用于小学 1~4 年级学生的思维能力测验, 该测验可以作为小学低年级思维能力的测查工具, 为儿童思维训练的效果提供评定依据, 也可以使教师对学生思维能力发展水平有一个较客观的认识, 是思维能力团体测验的有益探索, 并具有一定的教学指导意义。

## 2 测验编制过程

### 2.1 确定思维能力测验的分项目

思维能力是一个多层次、多侧面、多形态、多联系的复杂系统, 思维能力的各个成分相互联系、相互作用, 构成一个完整的结构。根据系统理论、

能力理论和思维理论, 我们提出了思维能力的结构是由思维内容、思维方法和思维品质构成的有机整体<sup>[4,5]</sup>, 对思维能力的测量必须同时考虑这三方面的因素。另外, 思维可以从不同的角度分为不同的类型。根据思维材料的不同, 思维主要可以分为形象思维和抽象思维。抽象思维是以概念为思维材料, 以概念、判断和推理的形式来反映事物的本质特征和内在联系的思维形式, 其思维方法主要包括: 分析与综合、抽象与概括、比较与分类、科学推理(包括归纳推理、演绎推理和类比推理)等。因此, 抽象思维能力的结构主要包括分析综合能力、归纳能力、演绎能力、抽象能力、概括能力、比较分类能力; 形象思维是以表象为思维材料而进行的思维, 其思维方法主要包括想象、联想、空间转换。因此, 形象思维能力的结构主要包括想象、联想、空间转换等<sup>[6]</sup>。根据上述对思维能力的理解, 考虑到小学生思维能力的发展水平, 本思维能力测验采用了图形、文字和数字三种丰富的材料, 包括 6 个分测验, 即比较分类(14)、归纳推理(12)、演绎推理(12)、空间认知(12)、类比推理(12)和抽象概括(13), 共 75 道题目, 为操作方便, 全部采用“四选一”的选择题形式。

### 2.2 测验题目的筛选

测验题目在山西省三所小学随机选取被试 242 人进行试测, 有效回收问卷为 231 份。利用区分度和鉴别指数两个指标, 对测题进行初步的项目分

收稿日期: 2007-11-8

\* 本研究得到教育部优秀青年教师资助计划 2003-110 项目的资助。

作者简介: 胡卫平, 男, 山西师范大学教育科学研究院教授, 博士生导师。Email: weipinghu@163.com.

析。利用各个题目与总测验的相关系数作为区分度的指标, 相关显著则说明可以代表要测量的内容, 相关不显著说明该项目缺乏鉴别力, 预测测验中相关不显著的题目有 11 个题目: 5, 15, 17, 23, 38, 67, 68, 69, 71, 73, 75。

鉴别度指数法, 即比较高分组 (27%) 和低分组 (27%) 的每一项目的通过率。鉴别度指数用高分组和低分组通过率的差值来表示, 即  $D=PH-PL$ ,  $D$  值越大, 项目的区分度越高。将被试按照测验的总分从高到低排序。找出高低分组上下 27% 处的临界分数, 46 分和 58 分为临界分数, 高分组和低分组的  $t$  检验结果显示, 删除在 0.05 水平上没有差异的项目: 2, 3, 4, 6, 7, 43, 52, 53, 55。

### 2.3 正式测验

思维能力测验初始题目一共 55 道题, 在三所小学进行施测, 发放问卷 816 份, 有效问卷 793 份, 缺失为 23 份。一年级 202 份 (男 103 人, 女

生 99 人), 二年级 224 份 (男 117 人, 女 107 人), 三年级 165 份 (男 81 人, 女 84 人), 四年级 192 份 (男 109 人, 女 83 人)。思维能力测验采用团体施测, 测验时间为 30 分钟, 每人发一份试卷和一张答题纸, 统一指导语, 统一收卷。

### 2.4 数据的统计处理

对测试的结果用 SPSS12.0 和 Liserl8.2 进行统计分析。

## 3 结果分析

### 3.1 项目分析

#### 3.1.1 难度分析

用项目的通过率来表示, 即项目回答正确的比率。本文把测验的难度分为 5 个等级,  $p<0.3$  (最难),  $0.3\sim0.49$  (较难),  $0.5\sim0.69$  (中等难度),  $0.7\sim0.89$  (较易),  $0.9\sim$  (最易), 各分测验与总测验的项目的难度等级分布见表 1 所示。

表 1 各分测验和总测验的项目难度分布

| 难度等级     | 比较分类                             | 归纳推理                             | 演绎推理                                | 空间认知                             | 类比推理                             | 抽象概括                             | 总测验                                |
|----------|----------------------------------|----------------------------------|-------------------------------------|----------------------------------|----------------------------------|----------------------------------|------------------------------------|
| $<0.3$   | 0 <sup>a</sup> (0 <sup>b</sup> ) | 0 <sup>a</sup> (0 <sup>b</sup> ) | 1 <sup>a</sup> (11.1 <sup>b</sup> ) | 0 <sup>a</sup> (0 <sup>b</sup> ) | 0 <sup>a</sup> (0 <sup>b</sup> ) | 0 <sup>a</sup> (0 <sup>b</sup> ) | 1 <sup>a</sup> (1.8 <sup>b</sup> ) |
| 0.3~0.49 | 2 (14.3)                         | 2 (22.2)                         | 4 (44.4)                            | 0 (0)                            | 2 (15.4)                         | 0 (0)                            | 10 (18.1)                          |
| 0.5~0.69 | 2 (14.3)                         | 3 (33.3)                         | 3 (33.3)                            | 3 (50)                           | 3 (23.1)                         | 0 (0)                            | 14 (25.4)                          |
| 0.7~0.89 | 9 (64.3)                         | 3 (33.3)                         | 1 (11.1)                            | 2 (33.3)                         | 7 (53.8)                         | 1 (25)                           | 23 (41.8)                          |
| 0.9~     | 1 (7.1)                          | 1 (11.1)                         | 0 (0)                               | 1 (16.7)                         | 1 (7.7)                          | 3 (75)                           | 7 (12.7)                           |

注: a 为该分测验对应的难度等级项目数, b 为该难度等级的项目个数占该分测验的百分比。

由表 1 可见, 比较分类测验中, 难度系数在 0.3~0.9 的项目占该分测验的 92.9%; 归纳推理测验中, 难度系数在 0.3~0.9 的项目占该分测验的 88.8%; 演绎推理测验中, 难度系数在 0.3~0.9 的项目占该分测验的 88.8%; 空间认知测验中, 难度系数在 0.3~0.9 的项目占该分测验的 83.3%; 类比推理测验中, 难度系数在 0.3~0.9 的项目占该分测验的 92.3%; 抽象概括测验中, 难度系数在 0.3~0.9 的项目占该分测验的 25%, 总测验中, 难度系数在 0.3~0.9 的项目占总测验的 85.3%。最难的题目和最易的题目较少, 整个测验难度属于中等难度偏易。各分测验对不同年级的通过率见表 2 所示。

从表 2 可见, 小学儿童在各个分测验和总测验上的通过率随着年级的升高而呈现升高的趋势。

表 2 不同年级学生在各个分测验上的通过率

| 分测验  | 一年级 (%) | 二年级 (%) | 三年级 (%) | 四年级 (%) |
|------|---------|---------|---------|---------|
| 比较分类 | 62      | 81      | 80      | 85      |
| 归纳推理 | 50      | 67      | 79      | 80      |
| 演绎推理 | 30      | 46      | 63      | 61      |
| 空间认知 | 56      | 72      | 75      | 91      |
| 类比推理 | 52      | 75      | 80      | 82      |
| 抽象概括 | 71      | 91      | 95      | 96      |
| 总测验  | 52      | 71      | 78      | 79      |

#### 3.1.2 区分度分析

采用鉴别力指数法来考察项目的区分度。鉴别度指数法, 即比较高分组 (27%) 和低分组 (27%) 的每一项目的通过率。鉴别度指数用高分组和低分组通过率的差值来表示, 即  $D=PH-PL$ ,  $D$

值越大,项目的区分度越高。按照美国心理测量专家 Ebel 提出的鉴别力指数评价项目性能的标准,将鉴别力指数分为<0.20 (区分度很差), 0.2~0.29

(区分度尚可), 0.30~0.39 (区分度较好), 0.4~0.49 (区分度很好) 四个等级<sup>[7]</sup>, 各分测验和总测验的区分度分布如表 3 所示。

表 3 各分测验的区分度等级分布

| 区分度等级    | 比较分类                   | 归纳推理                   | 演绎推理                | 空间认知                | 类比推理                | 抽象概括                 | 总测验                    |
|----------|------------------------|------------------------|---------------------|---------------------|---------------------|----------------------|------------------------|
| <0.2     | 6 <sup>a</sup> (42.8%) | 1 <sup>a</sup> (11.1%) | 0 <sup>a</sup> (0%) | 0 <sup>a</sup> (0%) | 0 <sup>a</sup> (0%) | 2 <sup>a</sup> (50%) | 9 <sup>a</sup> (16.4%) |
| 0.2~0.29 | 2 (14.2)               | 1 (11.1)               | 1 (11.1)            | 3 (50)              | 2 (15.4)            | 1 (25)               | 10 (18.2)              |
| 0.3~0.39 | 5 (35.7)               | 3 (33.3)               | 1 (11.1)            | 1 (16.7)            | 5 (38.5)            | 1 (25)               | 11 (20)                |
| 0.4~     | 1 (7.3)                | 4 (44.4)               | 7 (77.8)            | 2 (33.3)            | 6 (46.1)            | 0 (0)                | 20 (36.4)              |

注: a 为该分测验所对应的区分度等级的项目数, b 为该区分度等级的项目个数占该分测验的百分比。

比较分类测验中, D>0.2 的项目占该分测验的 57.2%; 归纳推理测验中, D>0.2 的项目占该分测验的 88.9%; 演绎推理、空间认知和类比推理的测验中, 所有项目的鉴别力指数都大于 2; 抽象概括测验中, D>0.2 的项目占该分测验的 50%; 总测验中, D>0.2 的项目占整个测验的 83.6%。按照鉴别力指数评价的等级标准, 整个测验的区分度较好。

### 3.2 信效度分析

#### 3.2.1 内部一致性和分半信度

量表的信度可以考察测验的内部一致性系数和分半信度。总测验的克隆巴赫 Cronbach  $\alpha$  系数为 0.892, 具有良好的信度。各个分测验的内部一致性系数和分半信度系数如表 4 所示。

表 4 内部一致性系数和分半信度系数

|             | 比较分类 | 归纳推理 | 演绎推理 | 空间认知 | 类比推理 | 抽象概括 | 总测验  |
|-------------|------|------|------|------|------|------|------|
| 分半信度        | 0.62 | 0.60 | 0.50 | 0.37 | 0.65 | 0.51 | 0.83 |
| $\alpha$ 系数 | 0.66 | 0.66 | 0.55 | 0.43 | 0.71 | 0.53 | 0.89 |

#### 3.2.2 各分量表和总量表的相关系数

结构效度可以用各个分量表和总量表的相关系数矩阵来考查, 见表 5 所示。

各个分量表之间的相关系数较低 (除空间认知外), 在 0.45~0.57, 但分量表和总量表的相关较高, 说明总量表具有良好的结构效度。

表 5 各个分量表之间的相关系数

|      | 比较分类   | 归纳推理   | 演绎推理   | 空间认知   | 类比推理   | 抽象概括   |
|------|--------|--------|--------|--------|--------|--------|
| 归纳推理 | 0.53** |        |        |        |        |        |
| 演绎推理 | 0.46** | 0.52** |        |        |        |        |
| 空间认知 | 0.65** | 0.62** | 0.53** |        |        |        |
| 类比推理 | 0.57** | 0.53** | 0.39** | 0.61** |        |        |
| 抽象概括 | 0.53** | 0.49** | 0.45** | 0.56** | 0.48** |        |
| 总测验  | 0.82** | 0.80** | 0.73** | 0.88** | 0.71** | 0.72** |

注: \*\*p<0.01。

#### 3.2.3 验证性因素分析

对思维能力的结构进行了验证性因素分析, 数据结果很好的拟和了模型。用 Lisrel 8.2 对测验的数据进行验证性因素分析, 参数估计如下图 1 所

示, 从验证性因素分析的拟合指数来看, 模型拟合的很好。各个拟合指数见表 6 所示。

从表 6 中, 可以看出模型拟合的非常好 (NFI、CFI、IFI、GFI 和 NNFI>0.95, RMSEA<0.08), 这

证明测验有很好的结构效度, 也说明这几种能力是思维能力的基本构成要素。

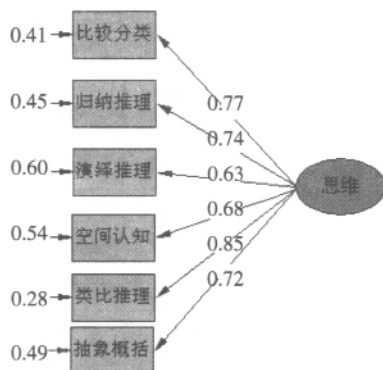


图1 思维能力测验验证性因素分析参数估计

表6 验证性因素模型拟合指数

| DF | $\chi^2$ | RMSEA | NFI  | NNFI | CFI  | IFI  | GFI  |
|----|----------|-------|------|------|------|------|------|
| 9  | 31.5     | 0.056 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 |

### 3.2.4 测验的年级特征

小学生思维的发展随年龄的增加而呈现上升趋势, 从测验的数据可以看到这种趋势, 结果见表7。

表7 小学思维能力测验的年级特征

| 分测验  | 一年级              | 二年级              | 三年级              | 四年级              |
|------|------------------|------------------|------------------|------------------|
| 比较分类 | 8.72 $\pm$ 2.23  | 11.40 $\pm$ 2.20 | 11.23 $\pm$ 1.67 | 11.92 $\pm$ 1.67 |
| 归纳推理 | 4.53 $\pm$ 1.79  | 6.09 $\pm$ 1.59  | 7.15 $\pm$ 1.56  | 7.23 $\pm$ 1.56  |
| 演绎推理 | 2.74 $\pm$ 1.31  | 4.15 $\pm$ 1.59  | 5.69 $\pm$ 1.60  | 5.48 $\pm$ 1.84  |
| 空间认知 | 3.34 $\pm$ 1.43  | 4.31 $\pm$ 1.22  | 4.47 $\pm$ 1.12  | 4.67 $\pm$ 1.13  |
| 类比推理 | 6.72 $\pm$ 2.28  | 9.78 $\pm$ 2.16  | 10.48 $\pm$ 1.90 | 10.64 $\pm$ 1.63 |
| 抽象概括 | 2.84 $\pm$ 1.02  | 3.59 $\pm$ 0.68  | 3.81 $\pm$ 0.48  | 3.82 $\pm$ 1.46  |
| 总分   | 28.91 $\pm$ 7.10 | 39.19 $\pm$ 7.35 | 42.84 $\pm$ 5.73 | 43.77 $\pm$ 5.07 |

从测验的原始数据来看, 分测验中, 归纳推理、空间认知、类比推理、抽象概括能力均随年级的升高呈现上升的趋势, 二年级学生的比较分类能力要比三年级高, 而三年级学生的演绎推理能力要比四年级高。从总测验来看, 思维能力测验的总分呈现上升趋势, 说明该测验有较高的效度。

### 3.2.5 测验和瑞文推理的相关

在进行思维能力测验的同时, 也实施了瑞文推理能力的测试。瑞文推理是目前使用最广泛的推理能力的测验, 本文用瑞文推理来作为本测验的效标关联指标, 结果见表8。

表8 测验与瑞文推理的相关

|    | 比较分类 | 归纳推理 | 演绎推理 | 空间认知 | 类比推理 | 抽象概括 | 总测验  |
|----|------|------|------|------|------|------|------|
| 瑞文 | 0.50 | 0.54 | 0.54 | 0.45 | 0.62 | 0.53 | 0.70 |

总测验和瑞文推理的相关系数为0.70, 除了空间认知外, 其他几个分测验与瑞文推理的相关都在0.50以上, 说明该测验跟瑞文推理具有较好的效标关联效度。

## 4 讨论

### 4.1 思维能力测验的编制结果分析

从本测验测试的山西省三所小学1~4年级793个学生的问卷分析来看, 整个测验的信效度基本达到测量学的要求。总测验中, 难度系数在0.3~0.9的项目占总测验的85.3%, 最难的题目和最易的题目较少。由于个别项目题目较易(抽象概括、比较分类), 总测验的难度中等偏易, 尚未达到正态分布, 这是需要改进的地方。

整个测验题目中, 有83.6%的题目具有较好的区

分度, 根据测量学的要求, 已经达到优良的水平<sup>[7]</sup>。测验的各个分项目的相关较低, 和总测验的相关较高, 相关系数在0.71~0.88之间。对6个分测验作了验证性因子分析, 数据表明模型拟合的很好(可参见表8), 思维能力测验的成绩随着年级的升高, 测验的总分也在增加, 具有明显的年级特征, 本测验编制的结构效度良好。总测验的分半信度为0.83, 内部一致性系数为0.89, 各个分项目(除空间认知)的分半信度和内部一致性系数较好, 分别为0.50~0.66, 0.53~0.71, 由于空间认知分测验题目少, 且只测查了其中的几个部分, 并且题目没有明显区分空间认知的几种能力, 而是综合起来考查, 将空间意识和空间转换结合起来, 将图形组合分解能力和图形特征抽象与概括结合起来, 所以这一部分还有待于进一步修正。由于测验采用了“四



选一”的选择题形式,无法排除被试有猜测的嫌疑,这也是选择题型无法避免的一点。

#### 4.2 小学生思维能力的结构及发展

思维能力是一种综合的能力,主要表现在比较分类能力、归纳推理能力、演绎推理能力、空间认知能力、类比推理能力和抽象概括能力。比较分类能力是小学生最重要的思维能力,是学生最先发展的思维能力<sup>[9]</sup>。推理是从已知信息得出推论和评价,这是一种认知性的任务。归纳推理是思维的主要认知形式和重要的思维能力,根据 Sloutsky 的标签属性模型,儿童特殊归纳推理的能力是以 7~8 岁为分水岭,也就是小学二年级学生的归纳推理能力是一个分界点<sup>[9]</sup>,演绎推理更多出现一种假言推理和三段式推理,学生在演绎推方面,不如归纳和类比推理<sup>[10]</sup>,这与本文测验编制的测量结果是一致的。抽象概括能力是探索和揭示事物的本质和规律的能力,本测验中主要是通过对于概念和种属关系来测试的,材料均为日常生活中常见的物体。空间认知能力是对思维能力的重要衡量,本测验主要是采用字符或者图形的表象转换、图形组合分解能力、空间意识<sup>[11]</sup>来测查的。

## 5 结论

自编的 1~4 年级思维能力测验的总测验的 Cronbach  $\alpha$  系数为 0.892,分半信度为 0.71,基本符合测量学的要求,与瑞文推理测验的皮尔逊相关为 0.673,具有较好的效标关联效度。各个分测验之间的相关系数较低,而与总测验的相关系数较

高,在 0.64~0.80 之间,具有较好的结构效度,经过分析,也具有适当的难度和区分度。验证性因素分析结果数据也很好的拟合了模型,本测验可以作为评估小学低年级学生思维能力的工具。

## 参 考 文 献

- 1 林崇德. 学习和发展. 北京: 北京师范大学出版社, 2003
- 2 张庆林, 陈敏. 小学儿童思维能力团体测验的初步研究. 西南师范大学学报 (自然科学版), 2000, 8: 481~486
- 3 白学军. 儿童思维发展水平测验的编制与测试结果. 心理发展与教育, 1994, 1: 22~25
- 4 胡卫平, 罗来辉. 论中学生科学思维能力的结构. 学科教育, 2001, 2: 27~31
- 5 胡卫平. 科学思维培育学. 北京: 科学教育出版社, 2004
- 6 尚凤祥. 思维能力的结构及思维品质的优化. 首都师范大学学报 (社科版), 1994, 4: 77~84
- 7 戴海崎, 张峰, 陈雪枫. 心理与教育测量. 广州: 暨南大学出版社, 2002
- 8 方富熹, 方格. 学前儿童分类能力的初步实验研究. 心理学报, 1986, 2: 157~164
- 9 Sloutsky V M, et al. How much does a shared name make things similar? Linguistic labels, similarity and the development of inductive inference. Child Development, 2001, 72: 1695~1710
- 10 陈安福. 中学生掌握假言判断的认知特点. 见: 刘范, 张增杰 (主编). 儿童认知发展与教育. 北京: 人民教育出版社, 1987
- 11 李洪玉, 林崇德. 中学生空间认知能力结构的研究. 心理科学, 2005, 2: 269~271

## THE PRELIMINARY DEVELOPMENT OF THINKING ABILITY SCALE FOR 1~4 GRADE ELEMENTARY SCHOOL STUDENTS

Hu Weiping, Li Haiyan, Qin Yaping

(Education Research Academy, Shanxi Normal University, Linfen 041000)

## Abstract

This study describes the development of thinking ability scale for 1~4 grade elementary students. The scale includes the items of characters, digital and graph test materials, which form 6 subscales. They are comparing and classification, the inductive reasoning, the deductive inference, the spatial cognition, the analogical reasoning and abstract-generalization. 793 1~4 grade students took part in the test. The results suggest that the scale has moderate difficulty, good discrimination, and high reliability and validity.

Key words Elementary school students, Thinking ability, Scale establishment.