



第二章 标准化测验

杨福义

博士 副教授

华东师范大学 学前教育与特殊教育学院

E-mail: fyang@spe.ecnu.edu.cn

Tel: 021-62232276



第一节 心理测量与测验的基本问题





一、测量的基本概念

1、测量最基本特征：将事物进行区分

2、测量的定义：

就其广义而言，测量是按照法则给事物指派数字。
[史蒂文斯（S.S.Stevens）]

3、测量包括三个要素：

1) 测量的对象——事物的属性和特征

2) 测量的规则——测量的规则或方法，即测量时给事物的属性指派数字的依据。

3) 测量的结果——描述事物属性的数字或符号。

4、测量的要素

1) 参照点

2) 单位





5、四种测量水平和测量量表

1) 量表的定义：在一个定有参照点和单位的连续体上把事物的属性表现出来，这个连续体称为量表。

2) 四种类型量表

类别量表 (nominal scale) : 区分性、同一性

等级量表 (ordinal scale) : 区分性、同一性、等级性

等距量表 (interval scale) : 相对零点

等比量表 (ratio scale) : 绝对零点

6、测量分为直接测量和间接测量





二、心理测量的基本概念

1、定义

心理测量就是根据法则对人的心理属性（行为）指派数字的过程。

2、特点

- 1) 间接性：心理测量均属于间接测量
- 2) 相对性：位置相对性
- 3) 多元性：对象复杂
- 4) 随机性：无关因素影响

3、心理测量水平：等级量表特征





三、 心理测验的基本概念

1、 定义：

测验是对一个行为样组进行测量的系统程序。

(F.G.Brown)

心理测验实质上是对行为样组的客观和标准化的测量。(A.Anastasi)

2、 心理测验三要素：

- 1) 行为样组：高度**代表性**
- 2) 标准化：测试的**一致性**
- 3) 客观性：衡量科学性的根本标志





3、心理测验客观性指标

1) 信度：同一群受测者在同一测验上多测量结果的一致**性**。

信度的种类：

重测信度

复本信度

内在一致性信度

2) 效度：一个测验所能测量的心理特质的**程度**。

效度是与一定的测量目的有关。

效度指的是某种程度，取值范围0-1之间。效度系数越接近

1，表明测验对所要测量的心理特质测得越准确。

3) 难度与辨别力

好的测题是鉴别力高且难度适宜的项目。





4、测验的种类

1) 根据测验的**对象**

认知测验——能力测验

人格测验——非能力测验

2) 根据测验的**人数**

个别测验

团体测验

3) 根据测验**材料**

语言或文字测验

非语言或操作性测验

4) 根据**编制的要求**不同

标准化测验

非标准化测验





5、测验的应用

- 1) 了解个别差异
- 2) 诊断、预测和评价
- 3) 甄选、分类和安置
- 4) 为心理辅导和心理咨询服务
- 5) 心理和教育科研的辅助手段





第二节 标准化测验的准备与实施





一、测验的选择

1、选择与测验活动目的相符的测验

分析与判断哪个测验最符合自己的测验目的

2、了解测试对象的受测条件

所选的测验应适用于特殊儿童

3、分析所选测验的特点

所选的测验必须具有良好的心理计量学性能

1) 题目有适当的难度和区分度

2) 测验结果具有很高的可靠性和有效性

3) 测验手册中提供标准答案、计分规则和适宜的常模

4) 施测、计分和分数的解释方法简单易行

心理测验使用者在对被试选择测验进行测试时，需依据**测验的目的、测验对象和所选测验的特点。**





二、测验人员具备的条件

1、知识与技能

心理测量理论

实际操作心理测验的技能和经验

2、职业道德

对自己的工作后果负责

认识自己能力的局限性

认识测验的局限性

注意对测验结果保密

不滥用测验





三、影响评估结果的因素

1、测验本身

- (1) 测题本身的代表性
- (2) 测题的数量
- (3) 测题的种类

2、测验实施

- (1) 测验情境
- (2) 主试的影响
- (3) 意外干扰





3、被试状态

- (1) 生理状况
- (2) 情绪状况
- (3) 练习因素





四、测验的准备与实施

（一）测验的准备

1. 主试的准备

2. 测验环境的准备

- 房间
- 座位安排
- 测验材料准备得当，置于被试看不见的地方
- 测试时只有主试和被试在场（主试助理一名）

3. 被试的准备

- 时间的选择
- 主试和被试建立和睦的关系，以期其良好合作，可以采取以下方式





(二) 测验的实施

1. 测验前的准备工作
2. 测题的选定
3. 如何记录
4. 如何对待被试的要求
5. 主试注意事项
6. 测验中可以安排短暂的休息
7. 测验结束后对被试表示感谢





第三节 标准化测验结果的解释





一、测验的评分

原始分数（raw score）：测验实施后，按照标准答案对受测者的反应逐项评分所得的分数为原始分数。

原始分数表示测验结果存在两大局限：

1) 原始分数意义不明确。各个测验或分测验的长度、满分及难度不同，同样的分数可能有不同的含义。不同的分数不能直接比较。

2) 原始分数的单位不等值。相同分数差距往往对应不同的真实差距。

2、导出分数(derived score):将原始分数转换到具有一定参照点和单位的数量连续体上的分数。





1、常用的概率抽样方法

- 1) 简单随机抽样
- 2) 等距抽样
- 3) 分层随机抽样
- 4) 整群抽样

2、除提供一般常模外，还应制定一些特殊常模
仅仅将特殊儿童与普通儿童相比较是不够的，有时还需了解特殊儿童在同类儿童中的相对位置。





3、常模

1) 以全体受测者或某个具有代表性的受测团体在测验上的分数分布为依据建立起来的，有一定参照点和单位的数量连续体叫做常模，有时也叫做常模量表。

解释测验结果的**参照标准**

2) 常模的制定：依据测试对象总体的平均成绩

3) 常模的可信度：取决样组的**代表性**和**可靠性**

2、常模团体：指的是具有某些重要特征的受测者群体，或者该群体的一个样本。





4、发展性常模和发展量表

1) 智龄 (MA)

1908年修订的比奈—西蒙量表中开始使用智龄来度量智力。智龄单位并不能保持恒等，智龄单位随年龄而减小。

2) 年级相当

年级当量比其他常模分数更容易产生误解。

比较同一被试在不同领域年级当量得分也可能导致误解。

有些测验，仅适用于一个很狭窄的年级范围（如3-5年级），分布的尾端是将曲线的两端延展，虽然延展到的年级水平被试可能实际上没有参加测验。

年级单位是不等的。





3) 顺序量表

这种量表上，分数可用相近的年龄水平来表示，同时还能对儿童的行为做质的描述。

4) 比率智商

存在的问题：

计算高年龄组儿童时应该用何实际年龄作为除数，尚无一定定论。

智龄不是一个等距单位，显然给求智商带来困难。

如果各年级组的标准差编制的不相等，则一个儿童所得到的智商在各年龄段不相同，给智商稳定性的解释带来麻烦。





4、组内常模和量表

几乎所有的**标准化测验**都提供某种形式的组内常模。

1) 百分等级

百分量表示等级（顺序）量表，无法适当将它加減乘除。

2) 标准分数

等距量表

2.1 标准Z分数

$$Z = \frac{X - \bar{X}}{s}$$





2.2 常态化的标准分数

通过对原始分数的非线性转化而来，与线性的Z分数不同，常态化的标准分数有一个近似正态曲线的分布。

2.3 Z分数的转化

小数与负数的存在是线性的和常态化的Z分数在计算和解释上有些不便。为此，将Z分数作线性变换。

$$Y=m+k(Z)$$

Y是转化后的分数，m和k为常数。所选择的m将为转化后新的分数分布的均数，而k则为标准差。





2.4T分数

最早T分数是由麦考尔根据12岁儿童团体来定义的，用来报告儿童在一份智力测验上的作业成绩。

$$T=50+10 (Z)$$

T分数的优点：1) 单位的等距 2) 可以对两个以上不同测验分数进行比较

T量表不能为未学过统计学的人掌握，但可根据常态曲线将T量表转化为百分比。





2.5 离差智商

美国心理学家韦克斯勒针对比率智商的缺陷而提出离差智商。

计算公式： $IQ=100+15z'=100+15(X-M)/SD$

从不同测验获得的离差智商只有当标准差相同或接近时才能比较。

缺点：由于常态化，使得对于智力极低者打分偏高；对于智力极高者打分偏低。





二、测验结果的报告

1、测验结果的综合分析

1) 分析心理测量的特点

2) 不根据一次测验结果下定论

3) 对测验结果的分析，从个体间比较和个体内差异两方面进行。

2、测试结果的解释和建议

