

# 2010 人力资源精品课程

## 北京大学

### 《人才测评学》全书

第一章 绪论

第三章 人事测量的基本类型

第五章 职业适应性测验

第七章 组织人事调查

第九章 测量的信度

第十一章 项目分析

第十三章 测量的实施与计分

第十五章 人事测量在现实中的应用

第二章 人员素质测评的历史

第四章 个性品质测验

第六章 能力测验

第八章 基于情境的测验

第十章 测量的效度

第十二章 测量工具的建立

第十四章 人事测量结果的解释

## 全国职业经理MBA双证班

**认证系列：**职业经理、人力资源总监、营销经理、品质经理、生产经理、物流经理、项目经理、企业培训师、酒店经理、市场总监、财务总监、行政总监、采购经理、营销策划师、企业管理咨询师、企业总经理等高级资格认证。

**颁发双证：**高级经理资格证书+MBA 高等教育研修结业证书（含 2 年全套学籍档案）

**证书说明：**证书全国通用、电子注册，是提干、求职、晋级、移民的有效依据

**学习期限：**3 个月（允许工作经验丰富学员提前毕业） **收费标准：**全部学费 **1280** 元

**咨询电话：**13684609885    0451- 88342620    **招生网站：**<http://www.mhjy.net>

**电子邮箱：**[xchy007@163.com](mailto:xchy007@163.com)    **颁证单位：**中国经济管理大学    **承办单位：**美华管理人才学校

## 全国招生 函授教育 颁发双证 权威有效



美华论坛  
[www.mhjy.net](http://www.mhjy.net)



职业经理 MBA 整套实战教程

千本好书 **免费** 下载    学校网址：[www.mhjy.net](http://www.mhjy.net)

# 全国Mini-MBA职业经理双证班



精品课程 权威双证 全国招生 请速充电

你可能准备跳槽或者求职, 却为缺少行业经验和专业证书而被用人单位百般挑惕!

你可能目前衣食无忧, 但随着年龄的增长和社会竞争压力的增大, 因为得不到专业的全新培训而失去竞争的机会和面临被淘汰的危机。

美华教育携手中国经济管理大学面向全国举办迷你 MBA 职业经理双证书班, 毕业颁发双证书。

## 招生专业及其颁发证书

认证项目	颁发双证	学费
全国《职业经理》MBA 高等教育双证书班	高级职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《人力资源总监》MBA 双证书班	高级人力资源总监职业经理资格证书+2 年制 MBA 高等教育研修证书	1280 元
全国《生产经理》MBA 高等教育双证班	高级生产管理职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《品质经理》MBA 高等教育双证班	高级品质管理职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《营销经理》MBA 高等教育双证班	高级营销经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《物流经理》MBA 高等教育双证班	高级物流管理职业经理资格证书+2 年制 MBA 高等教育结业证书	1280 元
全国《项目经理》MBA 高等教育双证班	高级项目管理职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《市场总监》MBA 高等教育双证书班	高级市场总监职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《酒店经理》MBA 高等教育双证班	高级酒店管理职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《企业培训师》MBA 高等教育双证班	企业培训师高级资格认证毕业证书+2 年制 MBA 高等教育研修证书	1280 元
全国《财务总监》MBA 高等教育双证班	高级财务总监职业经理资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《营销策划师》MBA 双证书班	高级营销策划师高级资格认证证书+2 年制 MBA 高等教育研修证书	1280 元
全国《企业总经理》MBA 高等教育双证班	全国企业总经理高级资格证书+2 年制 MBA 高等教育研修结业证书	1280 元
全国《行政总监》MBA 高等教育双证班	高级行政总监职业经理资格证书+2 年制 MBA 高等教育结业证书	1280 元
全国《采购经理》MBA 高等教育双证班	高级采购管理职业经理资格证书+2 年制 MBA 高等教育结业证书	1280 元
全国《医院管理》MBA 高等教育双证班	高级医院管理职业经理资格证书+2 年制 MBA 高等教育结业证书	1280 元
全国《企业管理咨询师》MBA 双证班	高级企业管理咨询师资格证书+2 年制 MBA 高等教育结业证书	1280 元



### 【授课方式】 全国招生、函授学习、权威双证

我校采用国际通用3结合的先进教育方式授课（远程函授+教学电子光盘自修+网络学院持续视频学习）



### 【颁发证书】 学员毕业后可以获取权威双证书与全套学员学籍档案

- 1、毕业后可以获取相应专业钢印《高级职业经理资格证书》;
- 2、毕业后可以获取2年制的《MBA研究生课程高等教育研修结业证书》;



### 【证书说明】

1. 证书加盖中国经济管理大学钢印和公章（学校官方网站电子注册查询、随证书带整套学籍档案）;
2. 毕业获取的证书与面授学员完全一致，无“函授”字样，与面授学员享有同等待遇，证书是学员求职、提干、晋级的有效证明；。



### 【学习期限】 3个月（允许有工作经验学员提前毕业，毕业获取证书后学校仍持续辅导2年）



### 【收费标准】 全部费用1280元（含教材光盘、认证辅导、注册证书、学籍注册等全部费用）

函授学习为你节省了大量的宝贵的学习时间以及昂贵的MBA导师的面授费用，是经理人首选的学习方式。



### 【招生对象】

- 1、对管理知识感兴趣，具有简单电脑操作能力（有2年以上相应工作经验者可以申请提前毕业）。
- 2、年龄在20—55岁之间的各界管理知识需求者均可报名学习。



### 【教程特点】

- 1、完全实战教材，注重企业实战管理方法与中国管理背景完美融合，关注学员实际执行能力的培养；
- 2、对学员采用1对1顾问式教学指导，确保学员顺利完成学业、胸有成竹的走向领导岗位；
- 3、互动学习（专家、顾问24小时接受在线咨询，第一时间回答学员的提问和咨询）



### 【考试说明】

1. 卷面考核：毕业试卷是一套完整的情景模拟试卷（与工作相关联的基础问卷）
2. 论文考核：毕业需要提交2000字的论文（学员不需要参加毕业论文答辩但论文中必修体现出5点独特的企业管理心得）
3. 综合心理测评等问卷。



### 【颁证单位】

中国经济管理大学经中华人民共和国香港特别行政区批准注册成立。目前中国经济管理大学课程涉及国际学位教育、国际职业教育等。学院教学方式灵活多样，注重人才的实际技能的培养，向学员传授先进的管理思想和实际工作技能，学院会永远遵循“科技兴国、严谨办学”的原则不断的向社会提供优秀的管理人才。



### 【承办单位】

美华管理人才学校是中国最早由教委批准成立的“工商管理MBA实战教育机构”之一，由资深MBA教育专家、教育协会常务理事徐传有教授担任学校理事长。迄今为止，已为社会培养各类“能力型”管理人才近10万余人，并为多家企业提供了整合策划和企业内训，连续13年被教委评选为《优秀成人教育学校》《甲级先进办学单位》。办学多年来，美华人独特的教学方法，先进的教学理念赢得了社会各界的高度赞誉和认可。



【咨询电话】13684609885 0451--88342620

【咨询教师】王海涛 郑毅

【学校网站】<http://www.mhjy.net>

【咨询邮箱】[xchy007@163.com](mailto:xchy007@163.com)



## 【报名须知】

- 1、报名登记表格下载后详细填写并发送邮件至 [xchy007@163.com](mailto:xchy007@163.com) (入学时不需要提交相片，毕业提交试卷同时邮寄4张2寸相片和一张身份证复印件即可)
- 2、交费后请及时电话通知招生办确认，以便于收费当日学校为你办理教材邮寄等入学手续。



## 【证书样本】(全国招生 函授学习 权威双证 请速充电)

(高级职业经理资格证书样本)

(两年制研究生课程高等教育结业证书样本)



## 【学费缴纳方式】可以选择以下任意一种方式缴纳学费

方式一	学校地址	<p>邮寄地址：哈尔滨市道外区南马路 120 号职工大学 109 室</p> <p>邮政编码：150020      收件人：王海涛</p>
方式二	学校帐号	<p>学校帐号：184080723702015</p> <p>账号户名：哈尔滨市道外区美华管理人才学校</p> <p>开户银行：哈尔滨银行龙江支行</p> <p>支付系统行号：313261018018</p>
方式三	交通银行 (太平洋卡)	<p>帐号：40551220360141505      户名：王海涛</p> <p>开户行：交通银行哈尔滨分行信用卡中心</p>
方式四	邮政储蓄 (存折)	<p>帐号：602610301201201234      户名：王海涛</p> <p>开户行：哈尔滨道外储蓄中心</p>
方式五	中国工商银行 (存折)	<p>帐号：3500016701101298023      户名：王海涛</p> <p>开户行：哈尔滨市道外区靖宇支行</p>
方式六	建设银行帐户 (存折)	<p>中国人民建设银行帐户(存折)： 1141449980130106399</p> <p>用户名：王海涛</p>
方式七	农业银行帐户 (卡号)	<p>农业银行帐户(卡号)： 6228480170232416918 用户名：王海涛</p> <p>农行卡开户银行：中国农业银行黑龙江分行营业部道外支行景阳支行</p>

可以选择任意一种方式缴纳学费，建议使用第五种方式(中国工商银行，比较方便快捷)收到学费的当天，学校就会用邮政特快的方式为你邮寄教材和考试问卷。

<http://www.mhgy.net>

## 第一章 绪论

### 第一节 人员素质测评的含义

#### 一、什么是人员素质测评

人员素质测评（人事测量）是心理测量技术在人力资源管理领域的应用。指的是以心理测量为基础，针对特定的人力资源管理目的，如招聘、安置、考核、晋升、培训等，采用科学的测评方法，对人的素质做出量值判断和价值判断的过程，为人力资源管理和开发提供依据。

（一）素质是个体完成一定活动（工作）和任务所具备的基本条件和基本特点。见图 1—1。

（二）测评是测量与评价的简称。

（三）测量是用数字或符号对人的特点进行描述，而不论其价值的大小。

（四）评价是依据定量描述或直觉经验来确定某种特点的价值。

（五）测量和评价都要依赖某种法则。

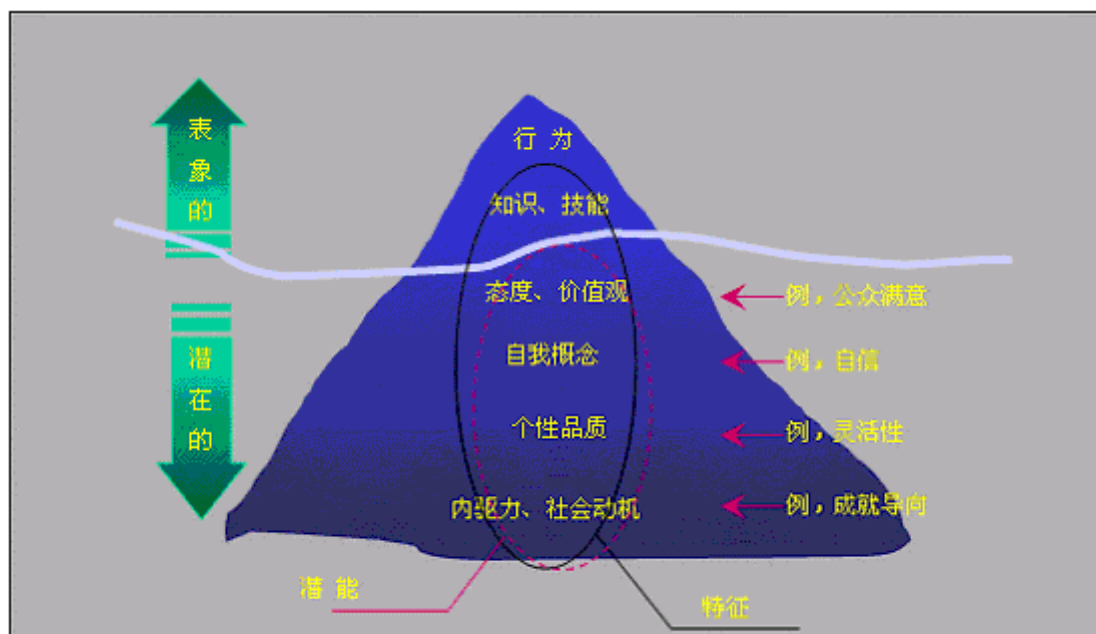


图 1-1 素质的冰山模型



## 第一章 绪论

### 第一节 人员素质测评的含义

#### 二、与人员素质测评相关的几个概念

（一）心理测量：通过科学、客观、标准的测量手段对人的特定素质进行测量、分析和评价。

（二）心理测验：是心理测量的一种具体手段。它是结合心理学或统计学方法，以评价特定个体在特定素质上相对于特定群体所处水平的手段。

从测量学的角度出发，心理测验由以下的五个部分构成：

##### 1. 行为样本

心理无法直接测量到，只有通过个体的行为和表现进行推测。在进行心理测量时，往往只能对经过科学选择的少数行为样本进行测查，借此推断个体的心理特征。由于实际上只考察了有限的行为样本，其结论是否可靠，取决于所抽选的行为样本能否很好地代表总体行为。

##### 2. 标准化

指测验的编制、实施、记分，以及测验分数的解释必须遵循严格统一的科学程序，保证对所有被测者来说，测验的内容、条件、记分过程、解释系统都相同。包括以下三个方面：

（1）测验题目的标准化：题目所引发的反应可以不同，但是题目内容必须相同。

（2）实施过程和记分的标准化：测验的环境相同，接受的指导语相同，记分方法相同。

（3）选用代表性的常模：常模是一组具有代表性的被试样本的测验成绩的分布结构，包含集中性和离散性。

集中性用平均数度量。离散性用标准差度量。所谓标准差是指一个数据序列各数据与平均数之差的平方和除以数据个数之商的平方根，通常记做 S。

$$S = \sqrt{\frac{\sum (X_i - M)^2}{N}}$$

##### 3. 难度的客观测量

测量的天花板效应：题目过于容易致使大部分个体得分普遍较高的现象。

测量的地板效应：题目过难致使大部分个体得分普遍较低的现象。

#### 4. 信度

测验结果的可靠性与稳定性。测验结果是否反应了被测者稳定的、一贯的真实特征。

（1）时间上的一致性

（2）内容上的一致性

（3）评分者的一致性

#### 5. 效度

测量的结果是否反映了所要测量的内容。

（1）效标关联效度

（2）内容效度

（3）构想效度

（三）心理测量、人员测评、心理测验的关系。见图 1—2。

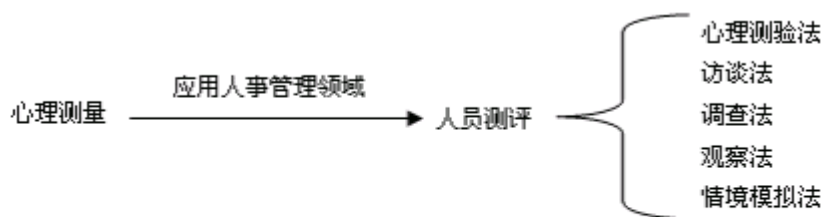


图1-2 心理测量、人员测评、心理测验的关系

## 第一章 绪论

### 第一节 人员素质测评的含义

#### 三、人员素质测评的假设系统

在这里，假设类似于数学中的公理，是指那些不言自明的命题。人员素质测评中的假设是这门学科发展的逻辑起点和前提。

具体的假设有三个：

假设一：个体差异

孔子：中人以上，可以语上也；中人以下，不可以语上也。

**坚强刚毅，可以立法，难与入傲，  
温柔和顺，可与循常，难与权毅，  
勇武雄悍，可与涉难，难与居约，  
谨慎或惧，可与保全，难与立节，  
凌厉刚正，可以持正，难与附众，  
能言善辩，可与范序，难与立约，  
宽弘博大，可以抚众，难以厉俗，  
偏激固执，可与守节，难以变通，  
深沉冷静，可与深虑，难以敏捷，  
性情善宽，可与赞善，难以矫违，  
好勇进取，可以进趋，难以持后，  
朴实直露，可与立信，难以消息。**

**——刘劭，《人物志》**

假设二：行为有因

人的行为受内部心理因素与外部环境制约： $B=f(P, E)$ 。P (Psychology) 代表内部心理因素，E (Environment) 代表外部环境。其中，心理因素主要由知、情、意组成。见图 1—3。



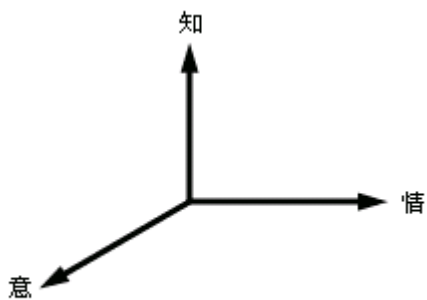


图1-3 人的心理特征结构

对于人力资源管理领域来讲，心理特征与管理行为之间也存在密切关系。见图 1—4。

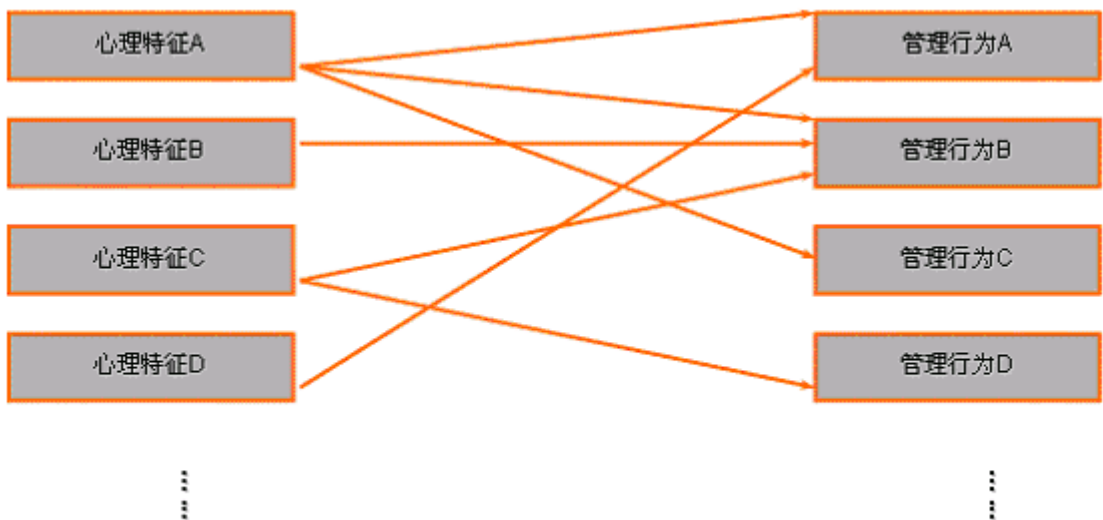


图 1-4 心理特征与管理行为的关系

假设三：人心可知

孟子：权，然后知轻重；度，然后知长短。物皆然，心为甚。

《史记》：尧使舜入山林川泽，暴风雷雨，舜行不迷。

## 第一章 绪论

### 第二节 人员素质测评的功能

人员素质测评通过科学方法对个体的行为和内在素质进行分析，是为人事决策提供可靠、客观的依据，为人事决策提供参考性建议，是人事决策的基本工具。从目前的管理实践来看，人员素质测评主要用于选拔、配置、考核、开发与晋升等几个方面。

## 一、选拔

随着社会分工的日益细化，以及经济社会的高速发展，个体与个体之间，组织与组织之间的竞争态势不断从广度与深度两个方面发展，工作本身对人的素质和心理适应性的要求越来越高。许多待遇优厚、工作舒适、社会地位高的职位常常具有众多的求职者，此时，单凭个人经验的选拔方法无法对人的心理素质进行科学准确的评估。在人力资源管理领域，人员素质测评可以用来对人才进行选拔。人员素质测评的运用可以使人事决策更为科学、准确，并可大大提高选拔的效率。

## 二、人员配置

不同的工作对人的素质要求有所不同，而每一个人的素质又具有不同的特点，这就要求在人员和工作之间选择最佳匹配。只有让每位员工做适合自身素质的工作，让每个岗位由具备该岗位素质要求的员工来担任，才能提高工作效率。通过人员素质测评，可以对个体的兴趣、人格、能力、技能等多方面进行分析，为实现人才的合理安置提供信息。

## 三、人员开发

不同的组织都将开发员工的潜能做为一项重要的任务。培训作为人力资源开发的重要组成部分，其内容越来越复杂，成本越来越大。运用人员素质测评可以判断员工各方面为工作所需要的素质，描述其素质结构，为制定有效的培训方案提供依据。

## 四、考核

众所周知，现代的人事考核已不仅仅局限于单纯的工作产出绩效考核，也越来越多地涉及工作中的行为、态度、胜任力等，这些内容的考核不同于传统的绩效考核，很难由直接的工作产出来表示，这就需要运用专门针对行为和内在品质进行量化描述、分析的心理测量技术。人员素质测评能够提供关于个体的行为的描述，形成对被测者的全面的评价，从而为人事考核提供依据。

## 五、晋升

晋升是人力资源管理的重要组成部分，是激励员工的重要手段。假如没有合理的晋升手段和晋升政策使得优秀的人员放到适合的岗位，同时，没有阻止无能力的人员到达不合适的岗位，那必然会增加企

业的人力资源管理成本，影响企业的快速发展。借助人员素质测评，可以使组织管理者对晋升的候选人进行考察，降低人事决策的风险。

## 第一章 绪论

### 第三节 人员素质测评的程序

#### 一、确定测量的目的和内容

根据不同的测量目的确定具体的测量内容是人事测量的第一步。测量内容应根据所选拔岗位的任职素质要求，通常可以工作分析、职务说明为依据，针对不同职务、不同岗位、不同企业特征的特殊要求来确定所需要的测量内容。

#### 二、确定测量的基本形式和测量工具

测量的形式和工具根据测量内容的不同而不同。测量形式和工具的确定也是非常重要的一步。不恰当的测量方法会使测量结果不能满足测量目的，甚至会导致收集到虚假信息，误导决策的制定。

一般自陈量表（即基于自我评价的问卷）的动机测验题目表面效度（即从题目表面是否容易看出出题人的意向和答案倾向）过高。应聘者容易表现出较高的社会赞许性，即题目本身的答案反映了一般社会价值倾向，应聘者容易表现出反应偏差，投其所好，故可能不适合于在招聘考核中采用时，就可以采用隐蔽性比较高的投射测验（如主题统觉测验）来对应试者的动机进行评定。

#### 三、测量的实施与数据采集

在测量的实施过程中，要注意做到客观化、标准化，保证收集到的测量结果能够公平、真实地反映应试者的状况。在收集测量资料的同时，要注意将实施测量的过程中相关的信息及可能对决策产生影响的细节记录下来，作为决策的辅助材料。另外，对测量结果造成影响的特殊因素也应进行记录。

通常，要求测量的现场环境要空气通畅、新鲜，照明充足，温度、湿度适宜，干净整洁，安静，没有外界干扰，每个被测者的桌椅应尽可能舒适，并有足够的空间，尤其避免多个被试者同时应试时相互影响、干扰。

#### 四、分析测量结果

对测量结果的分析通常包括对测量结果的计分、统计和解释。使用者只需按照测验手册的说明进行操作即可。

#### 五、根据分析作出决策或对决策的建议

决策与测量的目的联系紧密，以选拔为目的的测量，其决策内容为候选人名单；以安置为目的的测量，其决策内容为岗位与应聘者的匹配；以评价为目的的测量，其决策内容为对应试者素质的评价；以诊断为目的的测量，其决策内容为应试者的问题和特长或应试团体的状况和管理问题；以预测为目的的测量，其决策内容为应试者将来的绩效和工作表现。

#### 六、跟踪检验和反馈

在多数情况下，需要对测量结果及聘用结果进行跟踪，主要是根据工作绩效对测量结果和聘用进行检验，这就为此前的工作提供了重要的反馈，为测量取得经验性资料，为进一步矫正测量以达到更大的精确度提供依据。可以说，到这一阶段，才真正完成了一个人事测量作业系统。

## 第一章 绪论

### 第四节 人员素质测评的必要性与意义

一、人员素质测评是科学发展观在人力资源管理领域的具体体现。

（一）毛泽东：领导者的责任，一是出主意，二是用干部。

（二）邓小平：中国的事情能不能办好，社会主义和改革能不能坚持，经济能不能快一点发展起来，国家能不能长治久安，从一定意义上说，关键在人。

（三）江泽民：干部选拔工作要坚持走制度创新的路子，用科学的选人、用人制度来保障把人选准用准。

（四）胡锦涛：要建立以业绩为重点，由品德、知识、能力等要素构成的各类人才评价指标体系，建

立健全科学的社会化的人才评价机制。

二、个体之间存在巨大差异。

“面目仅半尺，竟无一人肖”；“人心不同，恰如其面”。

三、人事决策的困难性

没有什么决策比人事决策更难做出，后果会持续作用这么久。但是总的说来，领导者们所做的提升和人员调配决策并不理想，一般说来，平均成功率不大于 0.333：在多数情况下，三分之一的决策是正确的；三分之一有一定效果；三分之一彻底失败。

——Peter Drucker

人事决策是非常困难的，因为有多种因素可以影响人事决策。见图 1—5。

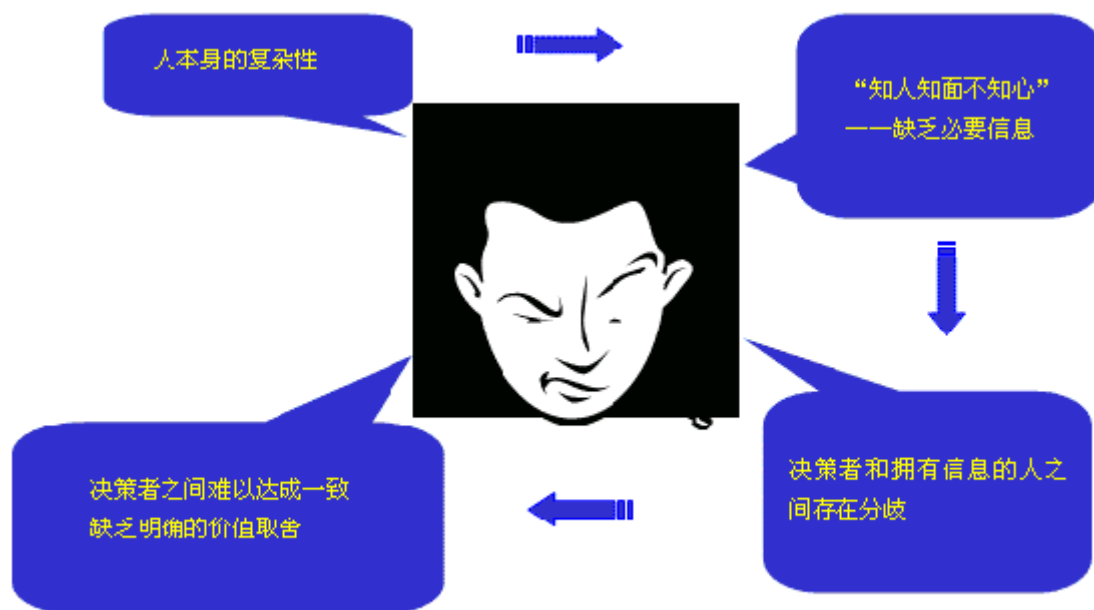


图 1-5 人事决策的影响因素

四、实现人职匹配是人员素质测评的核心目标，也是人力资源管理的一项重要功能。

## 第二章 人员素质测评的历史

### 第一节 我国古代人员素质测评的发展历史

#### 一、我国古代代表性的测评思想

在中国古代，思想家和政治家们对人才的识别进行了深入的探讨。形成了很多人才识别的思想和方法，称之为“识人法”或“识人术”。“识人法”或“识人术”的形成过程大致代表了人员素质测评研究在中国古代的发展过程。

孔子把人分为两类：“君子”和“小人”，并从思想心态和外在行为两方面提出其分类标准。思想心态方面，君子的标准用四个字来概括：德、义、刑、泰，小人则是土、利、惠、骄。

孟子特别重视对人的考察，他同齐宣王的一段对话中曾说过“权然后知轻重；度然后知长短；物皆然，心为甚”。意思是：人的心理能力和个性特征的差异是可以被测量出来的。这包含了对人的心理进行测量的数量化意识。

同样重视人才考察的还有墨子，墨子说：“听其言，迹其形，察其所能而慎予官，此谓‘事能’。故可使治国者，使治国，可使长官者，使长官，可使治邑者，使治邑。”（《墨子·尚贤中》）墨子主张通过观察人的言行，考察人的能力之后，根据能力大小再授予官职。这包含了人—职匹配的思想。

韩非子是我国古代历史上较早地发现考核的作用和意义的一位思想家，他主张考核的主要内容是工作能力。

至三国时期，刘邵写出了我国的第一部人才专著《人物志》。作者总结了我国历史上关于识人的思想与方法，对人才识别问题进行了系统研究，发表了精辟见解。其贡献主要表现在以下两个方面：

第一，根据德才关系，刘邵对人才进行了更精细的分类，他把人才概括为“三类”、“十二材”。“三类”即“兼德、兼材、偏材”，也就是德行高尚者、德才兼备者和才高德下者。这比孔子的“君子”与“小人”的分类更全面。“十二材”是指：（一）“国体”，兼有德、法、术三材，宜“三公之任”等；（二）“器能”，兼有三材，三材皆微，宜“冢宰之任”等；（三）“清节家”，德行高妙，容止可法，宜“师氏之任”等；（四）“法

家”，建法立制，强国富人，宜“司寇之任”等；（五）“术家”，思通道化，策谋奇妙，宜“三孤之任”等；（六）“臧否”，具有辨察是非之能，宜“刻削之政”（批评时政）；（七）“伎俩”，不能创思远图，而能受一官之任，有理烦纠邪之能，宜“司空之任”；（八）“智意”，不能创制垂则，而能遭变用权，有赞明计虑之能，宜“冢宰之佐”等；（九）“文章”，属文著述，宜“国史之任”等；（十）“儒学”，能够传圣人之业，不能干事施政，宜“安民之任”等；（十一）“口辩”，应对资接，宜“行人之任”等；（十二）“雄杰”，胆略绝众，材略过人，宜“将帅之任”等。

第二，传统儒家重德轻才，法家重才轻德，刘邵对这两家思想进行了融合，提出不同职位对德和才的要求各有侧重。同时，刘邵又列出了十二种类型的人才所各自适合担任的职位。刘邵认为，人材“能各有异”，“材能既殊，任政亦异”。这比墨子的人职匹配思想往前迈进了一大步。

## 二、我国古代代表性的测评方法

（一）周文王提出了六征法（见《大戴礼记·文王观人篇》）。六征法就是观察人的六个方面的特征。一曰观诚：观察验证一个人是否真诚；二曰考志：考察度量一个人的心志；三曰视中：考察一个人的内心；四曰观色：观察一个人的表情；五曰观隐：观察一个人隐藏伪托之处；六曰揆得：考察度量一个人的道德。

（二）孔子提出了“鉴人九法”。即“远使之而观其忠，近使之而观其敬，烦使之而观其能，卒然问焉而观其知，急与之期而观其信，委之以财而观其仁，告之以危而观其节，醉之以酒而观其侧，杂之以处而观其色。”

（三）诸葛亮提出了“七观法”。即“问之以是非，而观其志；穷之以辞辩，而观其变；咨之以计谋，而观其识；告知以祸难，而观其勇；醉之以酒，而观其性；临之以利，而观其廉；期之以事，而观其信。”七观法中，前四种是问答法，后三种是情境法。

（四）刘劭参照先秦时期思想家们的人材识别思想，提出了“八观五视”的方法。“八观”即“观其夺救，以明间杂；观其感变，以审常度；观其至质，以知其名；观其所由，以辩依似；观其爱敬，以知通塞；观其情机，以辩怨惑；观其所短，知其所长；观其聪明，以知所达。”“五视”即“居，视其所安；达，视其所举；富，视其所与；穷，视其所为；贫，视其所取。”



（五）我国古代智力考察工具。我国古代还出现了一些具体的智力考察工具，如七巧板和九连环等，这些工具的产生要比法国比奈的智力测验早数百年，但是和现代西方制作的智力测验量表相比，中国古代的这些工具并没有建立可靠的常模，因此，用七巧板和九连环等测量智力并不能得到像西方智力测验中的标准分数，难以判断被测者的智力相对于同类被测者所处的水平。而且，这些工具的出现并不是为了服务于人员素质测评，而仅仅是一种益智游戏。

### 三、我国古代的选人、用人制度

（一）禅让制：在“天下为公”的原始社会，流传着禅让制，通过让贤的方法来推选首领，这种可以认为是人材选拔制度的最早萌芽。据《史记·五帝本纪》记载，帝挚当了九年首领，政绩平平，而挚的部下尧聪明能干，建树颇多，于是人们便拥戴他，挚顺应民意，主动把首领的位置让给了尧，这就是历史上禅让制的开始。

（二）世官制：又称世卿制。世官制是夏、商、周三代最主要的人材选拔方式。禹的儿子启借助暴力夺取最高权力，建立夏朝，开创了中国古代王位世袭制的先河。这三代在世袭继承人的选择上有所不同：夏朝王位由儿子接任，偶尔也有传给兄弟的；商朝的王位继承在商代（盘庚当政时期）以前，一般实行“兄终弟制”，中期以后改为“父死子继”制。周朝则更进一步，确立了以嫡长子继承为主要内容的宗法制，“立嫡以长不以贤，立子以贵不以长”，这奠定了中国古代两千多年王位继承制的基本规范。

（三）察举制：察举制的基本内容是：首先由皇帝下诏规定察举的科目和要求，然后由丞相、诸侯王、公卿、刺史、郡国守相等中央和地方高级官吏，按照诏令指定的科目和要求，参照地方基层的评议考察人材，并向中央推荐。中央对所推荐的人材，经过一定的考核程序，由皇帝授予相应官职。

（四）九品中正制：九品中正制是在东汉末期察举制已经完全腐败的基础上，为适应新的社会政治发展的需要而建立起来的。所谓中正官是指由中央派遣的专门负责人材品评与选拔的各级人事官吏，是实施九品中正制的主体。其主要职责是“区别人物，第其高下”，即负责品评本州郡的人材，并向中央推荐。

“品”是中正官对本郡士人评出的等第，这与普通士人入仕的资格并且和入仕后所授职位密切相关，一般“品”高则所授职位高，“品”低则所授职位低。当时“品”分九等，即上上、上中、上下、中上、中中、中

下、下上、下中、下下。这就是“九品”。

（五）科举制：科举制创立于隋朝，并在唐朝进一步发展完善。科举制采用分科取士，历代科举制的科目设置不尽相同，总的来说，包括常科和制科两种。常科，如进士科、明经科、秀才科等，是科举考试制度中依规定要按时举行的固定的考试类型，一般每年举办一次，宋代以后一般每三年举办一次。常科考试通常实行三级或四级考试体制，不能任意跳跃。制科，是由皇帝为满足某种临时需要或选拔某种特殊人材而下诏举行的一种人事选拔方式。

## 第二章 人员素质测评的历史

### 第二节 现代人员素质测评技术的产生与发展

#### 一、现代人员素质测评技术的开始

##### （一）格林威治天文台事件

科学家最初发现人的心理的个体差异的重要性是起因于天文学上的一个事件。1796年，英国格林威治天文台的皇家天文学家N·马斯基林因为助手金内布鲁克观察星体通过的时间比自己迟0.8秒，认为他“师心自用，不依法行事”而将他辞退。此事在20年后受到另一天文学家贝塞尔的注意，他通过研究认为，这是一种不可避免的个人观察的误差。于是引起了学者们对个别差异的研究。但是，当时的心理学家们并不关心个别差异的测量。直到19世纪末时，心理物理学的发展才促进了对心理现象的量的测定。

##### （二）沈干的贡献

在西方一些国家，随着大工业革命的成功与工业化程度的加深，社会分工日益精细，导致专才训练与职业指导的需要，这也是促使测验发展的重要因素。特别是后来各种特殊能力倾向测验更是直接应实践的需要而产生的。最早关注人的智力因素的实践活动是从医学界发端的。大工业革命成功后，对劳动力的需求急剧增加，工厂大量采用童工。但是，许多地方官与工厂主订约，每雇20个童工，必须带一个低能者。在这种情况下，如何鉴别低能者成为现实的需求。第一个用科学方法研究低能者的是法国医生意太(Itard)。随后，意太的学生沈干(Sequin)继承其师的事业，用生理训练法进行了多年试验，并于1837年创办了第

一所专门教育智力落后儿童的学校。后来，其他一些欧美国家也陆续设立低能学校或在普通学校设特别班。

1864 年，沈干出版了《白痴：用生理学方法来诊断与治疗》一书，介绍了在感觉辨别力和运动控制力方面训练落后儿童的方法，其中的一些方法如形式板被后来的非言语智力测验所采用。

### （三）冯特（W. Wundt）的贡献——心理学的建立

1879 年，冯特（W. Wundt）在莱比锡大学设立第一所心理实验室。这一时期的实验心理学家长于生理学和物理学，主要目标是要发现人类行为的一般趋势。在研究中发现，对于同一刺激，各人的反应常常不同。研究者起初以为这是由于实验程序上的错误。经过长时间的实验才认识到，此种差异并非由于偶然的错误，而是由于个人能力上的真正差别。于是，引起了个别差异的研究。要研究个别差异必须有测量工具，由此便引起了测量运动。另一方面，早期的心理实验提出了严格控制观察条件的要求。例如，在一个测量反应时的实验中，给予被试者的指导语可能明显地提高或降低被试者的反应速度。再如，周围环境的照度和色调可能明显地改变视觉刺激物的形象。这表明，只有在标准状况下被试者所作的观察报告才是有价值的。这种程序的控制为测验的标准化打下了基础。

### （四）高尔顿的贡献——心理测验的尝试

首先倡导测验运动的是优生学创始人、英国生物学家和心理学家弗兰西斯·高尔顿爵士（Francis Galton）。高尔顿受 17 世纪英国教育家洛克“一切知识来之感官”的观点的影响，企图由各种感知觉辨别力的测量结果推估个人智力的高低。他在 1883 年出版的《人的能力研究》一书中说：“外部世界的信息是通过我们的感觉到达于我们大脑的。我们的感觉越敏锐，获得的信息便越多；获得的信息越多，我们的判断与思维便越有用武之地”。高尔顿还注意到，白痴对于热、冷、痛鉴别能力较低。这一观察结果使他进一步确信，感知觉辨别力“基本上是心智能力中最高的能力”。在测验中，高尔顿还是应用等级评定量表、问卷法以及自由联想法的先驱。另外，高尔顿还把统计方法应用于对个别差异资料的分析，他将以前数学家们所研究出来的统计技术改造为简单形式，使那些未经专门训练的调查者也能使用。他不但扩充了古特莱特（Guetelet）的百分位法，还创造了一种粗浅的计算相关系数的方法。后来高尔顿的学生卡尔·皮尔逊（Karl·Pearson）创立积差相关法，成为测验学者寸步不能离的工具。

### （五）比奈的贡献——心理测验的正式出现

美国心理学家宾特纳（R. Pintner）说：“在心理学史上，假使我们称冯特为实验心理学的鼻祖，我们不得不称比奈为心理智力测量的鼻祖。”

比奈（A. BiHe）1857 年生于法国尼斯市。比奈在青年时学习医学，却对心理学更有兴趣。1895 年他创办第一份法国心理学杂志《心理学年报》，同年他与亨利联名发表文章，批评当时流行的测验太偏重于感觉，过于集中在测量简单的、特定的能力方面。他们认为，测量比较复杂的功能，不必苛求精确度，因为这些功能的个别差异是较大的。1903 年，比奈在其著作《智力的实验研究》中提出，智力是广义的，应该包含一切高等心理过程，并且表现在推理、判断以及运用旧经验解决新问题的行为上。他以自己的两个女儿作为被试者，所用的测验有填字、图片解释等。有些后来也用于他的量表中。

1904 年法国教育部委派许多医学家、教育家与科学家组织一个委员会，专门研究公立学校中低能班的管理法。比奈亦是委员之一。他决心将测验的理论研究变为实际的应用，经过细心研究，次年与其助手西蒙（T. Simon）在《心理学年报》上发表一篇论文，题为《诊断异常儿童智力的新方法》，在这篇文章中介绍的就是第一个智力量表——比西量表。

1905 年的量表有 30 个由易到难排列的项目，可用来测量各种各样的能力，特别侧重于判断、理解、推理，亦即比奈所谓智力的基本组成部分。虽然这些测验也包括了感、知觉的内容，但其中言语部分所占的比例远较同时代的其他测验要大。不过该量表比较粗糙。从内容上看，该量表主要是对判断、理解和推理能力的测查，并不能涵盖智力的全部内容。从形式上看，这个量表无论是测验主持人的指导，还是被测者的回答或反应，主要都是以语言形式进行的，因而对语言能力的依赖程度较大，这对某些特殊儿童如聋哑儿童是不公平的。同时，这个量表不能准确地表示测验总分，只能以被测者完成题目的多少来大致确定其智力水平。也就是说，当时并没有一个确切的智力分数。为了弥补这种不足，1908 比奈对量表作了修订。采用智力年龄的方法计算成绩，即为每一个年龄（段）确定出难度最适宜的题目（即该年龄恰好有 60% 的被测者能完成的题目），用被测者所能完成的最高难度的题目所对应的年龄作为该被测者的智力年龄，以此对照其实足年龄（chronological age），来说明其智力的高低发展水平，并建立了常模，这是心理测验史

上的一个创新。

#### （六）心理测验的发展

比——西量表问世后，迅即传至世界各地。各种语言的版本纷纷出现。在使用中人们发现，沿用比——西量表的比率智商的概念会给记分和解释带来巨大问题。比如，一个3岁的孩子如果智力年龄仅为2岁，即智力（年龄）落后（实足年龄）1年，等到了6岁时，其智力年龄会是4岁，智力落后程度不再是1年而是2年，即其智力落后的程度不是保持固定的时间程度不变，而是按照一个比例，越来越大。从这时开始，人们才引入智商的概念，把（年龄）智商（Intelligence Quotient，缩写为IQ）定义为智力年龄除以实足年龄的商，再乘上100，以避免小数位并容易理解，写成公式就是：

$$IQ = \text{智力年龄} / \text{实足年龄} \times 100$$

这里的智力年龄是用经过测量检验能代表不同年龄智力水平的、有一定难度的题目来表示的。虽然对于较小的年龄，能给2个月甚至更小的年龄变化找出刚好代表其智力水平的题目，但到了一定的年龄如15岁以后，人们发现很难再找出一定数量的、在难度上能区分出不同年龄的题目。比如，很难找一些题目刚好是大多数18岁孩子能完成而大多数17岁孩子恰好不能完成。也就是说，实足年龄还在增长，但我们却找不到能代表其智力年龄增长的题目了。于是对以上公式又作出了修订，规定到了15岁以后，公式中的分母“实足年龄”不再变化，一律定为15岁。显然这个规定是比较人为化的。

到了20世纪中叶，由美国著名心理学家、斯坦福大学教授特曼（Terman）提出了离差智商的概念。

特曼认为：

1. 需要为不同年龄的人找不同的题目，而只要确定一组固定的题目，看不同年龄的人在这些题目上得分的分布，计算相应的平均数和标准差。
2. 有必要一定区分开不同的年龄，而是以实际测量结果为准，如果21岁和22岁的年龄组在得分上经统计检验没有显著性差异，就可以合并。
3. 智力水平的表示方法，可以将实际测量的分数以标准差为单位和本年龄组的平均分进行比较，也

就是说，以标准差为单位计算测验分数偏离平均数的方向和程度，这就是离差智商（deviation IQ），这个概念一直沿用至今。具体公式是：

$$IQ = 100 + 15 \times (X - M) / S$$

（ $X$  是实得原始分， $M$  为平均数， $S$  为标准差）

需要说明的是，离差智商不再是对智力的绝对意义上的度量，其分数本身并不能说明智力的实际高低，它只是一个相对值，表明一个人的智力水平相对于和他同类型的人群群体的平均水平所处的高低位置。

## 第二章 人员素质测评的历史

### 第二节 现代人员素质测评技术的产生与发展

#### 二、人员素质测评技术的发展

##### （一）人员素质测评技术在美国的发展

##### 1. 第一世界大战期间测评技术在美国的发展

1917 年，美国决定参加第一次世界大战。美国心理学会认为心理测验可以帮助军队对官兵进行选拔和分派。其中影响比较大的是在特曼的研究生奥蒂斯（A. S. Otis）所编团体智力测验的基础上发展出军用甲、乙两种测验，对 200 多万官兵进行了智力检查。战后此种测验经改造广泛用于民间，为教育与工商各界普遍采用。

除智力测验外，在一战期间被广泛应用的心理测验还有人格问卷。世界上第一个标准化的人格问卷——武德沃斯（Robert S. woodworth）编制的“个人资料调查表”（personal data sheet），就是为鉴别不能从事军队工作的神经症患者的需要而设计的。该问卷包括一些与常见的神经病症状有关的 116 个问题，由被测者以“是”和“否”的方式回答，对每个症状问题回答“是”，记 1 分；问卷总分是所报告的症状得分的总和。在一战结束前该调查表没来得及完成，战后，武德沃斯又编制了适用于一般人员和儿童的变式。

##### 2. 人员测评技术在美国工业界的发展

心理测量一诞生，就被工业界接受和运用，尤其是一些著名的大企业，开创性地在人事管理中引入心

理测量，为心理测量的普及起到了巨大推动作用。工业应用中最常见的心理测验有智力测验、人格测验、职业兴趣测验以及动机测验。

### （1）智力（能力）测验

在人事选拔中应用智力测验，既有考察智力水平的目的，也有考查智力结构的目的。一方面，不同的人智力水平不同，选择优智的人，可期望高绩效。另一方面，智力水平相近的人，其智力结构可能不同。有的人擅长言语理解、加工、表达，有的人擅长数字加工，有的人则擅长对形象的分析、加工。例如，在美国有一种测验叫做“分化能力性向测验”（Differential Aptitude Test，缩写为 DAT），分别从语言理解、语言推理、数学推理、抽象推理、空间推理、机械推理等六个方面检测人的智力水平，从而整体分析智力结构。这个测验的假定是：人的能力主要表现为这六个方面，而社会中的绝大多数职业对能力的要求也都可划分为这六类能力，并且，不同职业对不同方面能力的要求高低不同。另外一个重要的智力测验为威克斯勒（Wechsler）智力测验，包括成人版和儿童版；还有瑞文推理测验，这主要是以抽象图形推理的形式检测智力，其形式类似 DAT 中的抽象推理。另外，还有许多专门为工商企业开发的能力测验，它们大多专门针对某些具体的能力如数量分析能力测验、逻辑推理能力进行测验。另外还有一些针对特殊技能的测验，如打字测验，精确度与灵敏度测验，以及各种管理技能测验等。

### （2）人格测验

人格主要是指人所具有的与他人相区别的独特而稳定的思维方式和行为风格。人格同人的工作绩效和工作方式及习惯有关，概括来说，涉及以下几个方面：①人格会影响人在工作中对事物的理解。②人格会影响人处理事物的方法。③人格会影响人在工作中与他人相互沟通的方式。④人格会影响人独特的表现方式。因此，不同的人格会给工作带来不同的结果。这也就要求人事诊断和安置作出相应的考虑并采取相应的方法。心理学家研制出了种类相当多的人格测验，其中最著名的有：明尼苏达多项人格问卷（MMPI）、卡特尔 16 人格因素问卷（16PF）、加州人格问卷（CPI）、梅耶—布里基斯人格特质问卷（MBTI）、DISC 人格测验（DISC）等。

### （3）职业兴趣测验



职业兴趣也是人事选拔经常参考的一种心理测验。大量研究揭示，不同的人的工作生活兴趣可以按照对人、概念、材料这三大基本内容要素分类，而社会上的所有职业、工作也是围绕这三大要素展开的。因此，在工作所规定的核心要素与人所感兴趣的内容之间，恰好形成对应。这就意味着，要在工作与人的兴趣之间进行匹配。历史上最著名的职业兴趣测验有斯特朗坎培尔（Strong&Campbell）职业兴趣测验和霍兰德（Holland）职业兴趣问卷。他们都通过分析，确定职业兴趣分为六大类：社交型、艺术型、研究型、技能型、事务型、经营型。不难看出，通过鉴别人的职业兴趣，可以对人事选拔、安置提供重要的参考依据。

#### （4）动机测验

所谓动机（motivation）是指由特定需要引起的，欲满足该种需要的特殊心理状态和意愿。从以往的研究和测验来看，大致分为两个思路：一个是探讨人的一般生活动机；一个是专门探讨人的工作动机。所谓一般生活动机（general motivation）是指人们广泛的生活领域中具有普遍性的需求所导致的动机。探讨一般生活动机的主要理论有马斯洛（Maslow）的需要层次理论，赫兹伯格（Hertzberg）的保健——激励理论，奥德佛尔（Alderfer）的 ERG 理论（ERG 即生存、关系、成长）。这三种理论都试图分析人类基本需要的种类，从而对需要的内容作出说明，因此都属于需要的理论。而它们同时又被看做是动机理论，因为动机总是由需要而来的，解释需要也就解释了动机的来源。所谓工作动机（work motivation）是具体指驱使人们工作的原因。在这方面最有影响的当属哈佛大学著名心理学家麦克里兰（McClelland）的工作，他提出了著名的三重需要理论，认为人们的工作动机可以细分为三种：①成就动机，即寻求获得成功；②权力动机，即寻求获得、保持和运用对他人的影响和支配；③亲和动机，即寻求与多数人群保持密切的关系。

由于心理测量有如此重要的功能，因而在欧美企业中得到了普遍欢迎和大量应用。美国电话电报公司（AT&T）早在 20 世纪 30 年代就启用评价中心技术，采纳了许多人事测量的方法用于考查自己的管理者，并取得了相当的成功，为采用人事测量技术预测管理者的绩效和未来职业发展积累了重要的资料和经验。

## （二）现代人员测评技术在我国的发展

### 1. 蔡元培的贡献——1917 年北大成立心理学实验室

蔡元培先生曾经在德国莱比锡大学师从冯特教授，学过实验心理学。回国后，他在 1917 年于北京大

学成立了第一个心理学实验室，标志着心理学在中国的诞生。

## 2. 心理测验在中国的引入与传播

1916 年樊炳清把比奈——西蒙量表介绍到中国。随后费培杰、鲁志伟、吴天敏等人都对心理测验在中国的推广与应用做出了贡献。

## 3. 新中国建立前，心理测量技术在工业领域的应用。

## 4. 1979 年以后人事测量工作开始迅速发展。

1979 年，北京大学的吴天敏教授对中国的比奈量表进行了第三次修订。在随后的 20 年内，大批的组织与管理心理学家开始着手心理测量在人事管理中的应用，政府、企业都越来越重视人员素质测评技术在现代人事管理中的应用，大批专业机构大量涌现。

### （三）人员测评技术的最新发展趋势

#### 1. 人事管理领域的测验大量涌现。

#### 2. 面试技术在人才选拔中广泛应用。

#### 3. 人格测验成为人才甄选程序的必要过程。

#### 4. 针对具体要求的测验设计、组合越来越重要。



**职业经理 MBA 整套实战教程**

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第三章 人事测量的基本类型

### 第一节 标准化的纸笔测验

#### 一、什么是标准化的纸笔测验

在人事测量中，标准化的纸笔测验的应用最为广泛。顾名思义，纸笔测验（paper-pencil test）就是只

用纸和笔就能进行的测验。这类测验在实施过程中，一般不需要借助其他工具和手段。标准化的纸笔测验一般有事前确定好的测验题目和答卷以及详细的答题说明，测验题目往往以客观题居多，但也有不少是主观自陈评价题。有的测验限定时间，有的则不限定时间。被试的任务一般很简单，只需按照测验的指示语回答问题即可。一个标准的纸笔测验系统还包括客观的计分系统、解释系统、良好的常模以及值得信服的信度、效度和项目分析数据。

大多数智力测验、人格测验、成就测验、能力倾向测验等，都采用纸笔测验的形式。卡特尔 16 种人格因素测验、生活特性问卷、数量分析能力测验都是属于这种类型的测验。

## 二、纸笔测验举例

### （一）艾森克测验的题目（部分题目）

## 艾森克人格问卷

**方法：**在这张问卷上有 45 个问题，请依次回答这些问题，回答不要写字，只要在问题后的“是”或“否”上打勾或画圈即可。

这些问题要求你按自己的实际情况回答，不要去猜测怎样才是正确的回答。因为这里不存在正确或错误的回答，也没有捉弄人的问题，将问题的意思看懂了请尽快回答，不要花很多时间去想。

每一条都要回答。问卷无时间限制，但不要拖延太长，也不要未看清问题便回答。懂了吗？不懂可以。开始！

1. 你是否有许多不同的业余爱好？
2. 你的心境是否常有起伏？
3. 你是否健谈？
4. 你曾无缘无故觉得“真是难受”吗？
5. 你是否比较活跃？
6. 你是否常常为自己不该作所作的事，不该说或说了的话而紧张吗？
7. 你喜欢跟长辈争辩吗？
8. 通常你能在热闹聚会上尽情玩乐吗？
9. 你容易激动吗？
10. 你是一个容易伤感情的人吗？
11. 在社交场合你是否总不愿靠头角？

## （二）分数解释

### 1. 算原始分数

**表 3-1 艾森克量表的原始分对照表**

内外倾项目		情绪稳定性项目	
1	25	2	26
3	27	4	29
5	28	6	30
7	31	9	32
8	33	10	34
-11	35	12	36
13	40	14	37
-15	42	16	38
17	44	18	39
19		20	41
21		22	43
-23		24	45

注：有“-”号的回答“否”记分。

2. 将原始分转化为标准 T 分数

**表 3-2 艾森克量表的标准分对照表**

男性			女性		
T 分数	内、外倾	情绪稳定性	内、外倾	情绪稳定性	T 分数
30	1-2	1-3	1-2	1-4	30
35	3-4	4-5	2-4	5-7	35
40	5-6	6-8	5-7	8-9	40
45	7-9	9-10	8-9	10-11	45
50	10-11	11-12	10-11	12-13	50
55	12-13	13-14	12-13	14-15	55
60	14-15	15-17	14-16	16-18	60
65	16-17	18-19	17-18	19-20	65
70	18-20	20-21	19-20	21-22	70

3. 对分数的解释

对于内外倾：

50 分为平均分，低于 43 分为比较内向，低于 38 分为非常内向，高于 56 分为比较外向，高于 62 分为非常外向。

对于情绪稳定性：

50 分为平均分，低于 43 分为比较稳定，低于 38 分为非常稳定，高于 56 分为比较不稳定，高于 62 分为非常不稳定。

### 三、对标准化纸笔类测验的评价

#### （一）纸笔测验的优点

1. 方便性。首先，这种测验很容易实施，测验一般有详细的实施说明。一般就单一的测验而言，一位没有受过任何心理测量训练的主试，可以在很短的时间内学会如何操作测验的施测过程。这就使一些非专业人士也可以很好地使用这些测验。

2. 经济性。这类测验通常可以团体施测，可以节约大量的精力和时间，在较短的时间内获得被试者的大量信息。

3. 客观性。纸笔类测验较为客观，它往往有标准化的实施说明、计分系统和解释系统。测验结果受测验实施者和计分人员的主观因素的影响较小，可以保证在公平的前提下进行测验，应试者比较容易接受和信服。

#### （二）纸笔测验的缺点

1. 受测验的形式所制约，它无法对被测者实际的行为表现进行测量，如言语表达能力、操作能力等等。

2. 纸笔测验的实施较为程式化，我们只能收集到测验中所考察的信息，而对于测验外的信息我们一无所知。

3. 纸笔测验中有标准答案的测验如能力测验，并不能完全避免考试技巧和猜测因素的影响，而那些没有标准答案的主观自陈形式的测验如人格测验，也往往无法完全解决应试者掩饰自己的真实情况的问题。

## 第三章 人事测量的基本类型

### 第二节 投射测验

## 一、什么是投射测验

（一）投射最早出现于弗洛伊德的心理学理论。投射技术一词由富兰克（L. Frank）于 1939 年首先明确提出。

（二）投射测验（projection test）要求被试对一些模棱两可或模糊不清、结构不明确的刺激作出描述或反应，通过对这些反应的分析来推断被试的内在心理特点。其逻辑假定是，人们对外在事物的看法实际上反映出其内在的真实状态或特征。例如，设计一幅没有任何含义、模糊不清的墨迹图，要求被试看了墨迹图之后展开联想，回答主试提出的问题。

### （三）投射测验的特点

1. 隐蔽性。这是指测验的目的的隐蔽性，被试一般不可能知道测验的真实目的，也不知道对自己的反应将作何种心理学解释，减少了被试伪装自己的可能性。

2. 非结构性。这是指投射测验使用非结构化任务作为测验材料，即允许被试产生各种各样不受限制的反应。为了促使被试充分想象，投射测验一般只有简短的指示语，测验材料也是含义模糊、模棱两可。由于测验材料的模糊性，被试的反应较少受到情境线索和他人观点的影响，往往会表现出被试的真实的内在感受、需要、个性、情绪、动机、冲突、防御等心理内容。采用投射法可以测试出被试人格的更真实的面貌。

3. 整体性。这是指测验关注的是对人的总体评估，而不是针对单个特质的测量。被试的任何反应都可能影响评估结论，在对投射测验进行解释时要注意它的整体性特征。

## 第三章 人事测量的基本类型

### 第二节 投射测验

#### 二、投射测验的分类及实例

林德塞（G.Lindzey）按照测验所引起的反应性质，将投射测验分为五类：

##### （一）联想法投射测验



测验任务为报告因某种刺激（如字词、墨迹、图画）所引起的联想（通常是最先引发的联想）。

### **例 1：罗夏克墨迹测验**

**测试材料：**十张墨迹图，其中五张是黑白的（见图 3-1），两张有黑色和红色墨迹，另外三张为淡彩色墨迹图。

**测验任务：**看完图片后对图画内容进行自由联想，主试通常问：“请告诉我你在图片中看到了什么？或者想起了什么？”回答后主试会将图片再展示出来，询问是图片的哪个部分，哪些因素致使被试作出以上联想的。



图 3-1 墨迹测验示例

### **例 2：主题统觉测验**

**测试材料：**三十一张图片（见图 3-2），其中有一张是空白图片，其他图片中画有主题不明确的一些图画，大多数图画中包含人物。

**测验任务：**主试每次给被试看一张图片，要求被试根据图片中的图画讲一个故事，鼓励被试尽量发挥想象力，说出这个故事为什么会发生，会有什么结果等内容。被试所讲述的故事越详细越好。



图 3-2 主题统觉测验示例

## （二）构造法投射测验

测验任务为让被试编造或创造一些作品（如故事、图画等）。

### 例：绘人测验

**测验任务：**要求被试在一张白纸上用铅笔任意画一个人。画完后，再要求被试画一个与前者性别相反的人。最后，主试可以通过面谈的方式向被试了解他所画人物的年龄、职业、爱好、家庭、社交等信息。最后测验者对被试的作品进行分析。

## （三）完成法投射测验

测验任务为让被试对一项不完整的材料进行填补使其完整，如语句完成法。

### 例：罗特未完成的句子量表

**测验任务：**要求被试将一系列句子补充成完整的句子。通过被试的反应可以对被试的家庭态度、社会与性态度、一般态度、品德态度进行解释。

## （四）选择或排列法投射测验

测试任务为让被试将一些刺激按照某种原则进行选择或予以排列。例如，可以让被试将一些描述人格的词按照其好恶程度或适宜程度排序。从顺序中可以分析出被试的人格。

## （五）表露法投射测验

测试任务为被试利用为某种媒介自由地表露自己的心理状态，例如，可以通过书写、谈论、歌唱、绘画等形式让被试自由表达，从中分析其人格。

### 第三章 人事测量的基本类型

#### 第二节 投射测验

##### 三、对投射类测验的评价

通过投射技术可以使被试不愿表现的个性特征、内在冲突和态度更容易地表达出来，因而在对人格结构、内容的深度分析上有独特的功能。投射技术在临床领域有一定的应用前景。但是，对于投射测验一直存在相当多的批评：

（一）由于投射测验结果的分析一般是凭分析者的经验的主观推断，其科学性有待进一步考察。

（二）投射测验在计分和解释上相对缺乏客观标准，人为性较强，不同的测验者对同一测验结果的解释往往不同，并且，投射测验的重测信度也很低。

（三）投射技术是否能真正避免防御性反应的干扰，在研究上并未得出一致结论。

（四）投射测验一般为个体测验，不仅测验时间长，分析结果所需要的时间也很长，实施起来耗费精力。

（五）投射测验对主试和评分者的要求很高，一般只能由经验丰富、有专业背景的人担当。这种局限使一般的人事管理人员无法直接应用投射测验，测验的推广受到影响。

（六）对投射测验结果的评价带有浓重的主观色彩，不能满足人事测验的公平性原则。

虽然投射技术面临这么多的批评，心理测量学家们并没有放弃在这方面的努力，他们一直致力于使投射测验更加客观化和标准化。例如，经过改进的主题统觉测验利用内容分析的方法评分，对个体的成就动机、亲和动机、权力动机进行评估，使该测验具备了一定的客观性。相信通过不断改进，投射测验将能够在各种应用领域发挥其独特的作用。

### 第三章 人事测量的基本类型

### 第三节 行为模拟与观察类测验

#### 一、行为观察法

行为模拟与观察类测量的技术核心是行为观察法（behavioral observation），它是通过安排一定的情境，在其中观察特定个体（或群体）的特定行为，从中分析所要考查的内在素质或特征。行为观察法可以分为自然观察法、设计观察法和自我观察法。

##### （一）自然观察法

自然观察法（natural observation）是观察者在真实的生活或工作情境中对个体的行为进行直接观察的方法。在自然观察中，观察者不应该对情境作任何干预和改变，被观察者也不应该意识到自己正在被观察这一事实。观察者在观察的过程中对被观察者的行为进行详细的记录，包括利用各种观察、记录设备，如摄录像机，事后根据记录对被观察个体的行为进行分析和评估。

管理者对其下属的工作表现的观察和记录，可以算是自然观察的一种形式，只不过通常不够客观和科学。在企业中，可以尝试在考核中引进客观的自然观察方法。人事经理可以在考核前任意选定一个工作时段，在不被员工发觉的前提下对员工的工作行为进行观察，并作客观详细的记录（例如，记录员工接待客户时使用礼貌用语的频次），该记录资料可以作为考核的依据。

自然观察的内容真实性很高，往往能够反映被观察者的实际情况。但在观察时要注意观察时间的选取，因为有些行为并不一定经常出现，例如接待客户的礼貌行为在没有客户来访时并不会出现，如果把观察时间选定在客户通常很少来访的时候，则观察结果会受影响。另外，要保证被观察者所处的情境中不存在影响受观察行为的因素，例如不应选择被观察员工身体不适时进行观察，也不应选择被观察员工的上级在场的情况下进行观察。

##### （二）设计观察法

在需要对真实生活中不易随时观察到的行为进行评估时，可以采用设计观察的方法。例如，我们通常会想知道一个应聘者在紧张压力环境下的表现，但我们一般很难有机会在自然状态下遇到这种观察机会。这就需要采用设计观察法（designed situation observation），即在人为设计的环境中观察特定的行为或反应。

在人事测量领域经常用到的设计观察方法有情境压力测验和模拟情境测验两种。

### 1. 情境压力测验

情境压力测验（situational stress test）由主试向被观察者布置一定任务和作业，借以观察个体完成任务的行为。工作样本测验、无领导小组讨论都可算做情境压力测验。

在工作样本测验（working sample test）中，主试通常向被试布置一项工作任务，要求被试在一定时间内完成，观察者对被试完成任务的行为过程和行为结果进行观察和评估。例如，在招聘打字员时，考官通常让应聘者现场录入一篇文章或听打一封口授信函，对应聘者的打字速度、错误率进行评估。在招收演员时也常常采用工作样本的方法，例如让应聘者现场表演一个小品，以观察其演技及应对能力。

无领导小组讨论（leaderless group discussion）是安排一组互不相识的应聘者（通常为6—8人）组成一个临时任务小组，并不指定任务负责人，请大家就给定的任务（讨论题）进行自由讨论，并拿出小组决策（讨论）意见。这是一种目前很受欢迎的人事测量方法。在无领导小组讨论情境中，主试对每个被试在讨论中的表现进行观察（通常要使用专门的摄录像设备），但不参与到讨论中去。由于没有安排领导，而讨论任务往往又需要有人来对讨论的进程进行引导，对各类意见进行调节和总结，并最终导向一致决策，因此具有领导才能的个体往往可以在这种情境中脱颖而出。无领导小组讨论可以对三方面的内容进行测量：

- （1）个性特点：自信心、挑战性、自我表现的主动性、主动寻求他人同意和赞赏的倾向等。
- （2）社交能力：言语表达的流畅性、组织性和逻辑性，对他人的说服力，倾听的技巧等。
- （3）领导意识和领导能力：主动扮演领导角色的倾向、对进程的控制能力、意见综合能力等。

### 2. 模拟情境测验

模拟情境测验（situation simulation test）是指通过模拟一个尽可能接近真实工作情境的环境，要求被试完成某项任务，对被试的完成任务的行为过程及行为结果进行观察、评估。模拟情境测验与情境压力测验的唯一区别就是在模拟情境测验中被试处在一个假定的情境中，这个假定情境可以根据测验的需要不断变换，以充分考察被试在不同情境中的表现。角色扮演测验就是一个典型的模拟情境测验。

角色扮演测验（role taking test）是通过赋予被试一个假定的角色，要求按照角色的要求表现自己的行为，观察记录并评价角色扮演的行为，评价角色接近程度或胜任力。例如，在招聘销售员时，考官往往要求应聘者扮演销售人员，自己扮演客户，假定考场就是销售现场，要求应聘者向自己推销某一产品，借以考察应聘者的销售能力。在测验进行过程中，考官可以根据需要不断提出各种刁难性问题，使考生的能力可以得到充分体现。角色扮演测验的操作成本高，对考官的要求也很高，但是由于它对人员的评估十分有效，不少企业都很乐于采用这种方法。

### （三）自我观察法

自我观察法（self observation）是由被试自己对自己的行为进行观察，并记录自己行为的方法。自我观察有利于被试对自己行为的监控。企业中常采用的“工作日志”就是自我观察的一种应用。对工作行为的自我观察有利于观察者对自己的工作进行总结和改进。另外，自我观察法也可以用于对员工的考核。有些企业要求处于试用期的员工对自己每天的工作内容进行详细的记录。在试用期结束时，这些记录成为考察员工工作能力的重要依据。需要注意的是，采用自我观察法对员工进行考核，必须建立在员工如实反映情况的前提上。

## 二、对行为模拟与观察类测量的评价

### （一）行为模拟与观察类测验的缺点

1. 操作困难。这类测量对主试的要求很高，通常需要富有经验的主试来操作。
2. 成本较高。这类测量通常只能逐个进行，测量的成本很高。
3. 分析结果具有很大主观性。

### （二）行为模拟与观察类测验的优点

1. 通过行为测量可以观测到被试在具体情境下的真实行为，比对行为的自我评估，其结果更容易令人信服。
2. 通过对情境的操作，可以使测试情境与将来的工作情境尽可能相似，所测量出的结果具有很好的预测效果。

3. 在行为测量中，被试的个人特点和能力可以得到充分的展示。通过对情境的设计，考官可以成功地观测到许多他所想要考察的行为和能力。

### 第三章 人事测量的基本类型

#### 第四节 基于工作情境的综合类测验

##### 一、公文筐测验

公文筐测验（in-tray test）是让被试在所安排的假想的情境中扮演某种管理者的角色，对事先设计的一系列文件进行处理，进而针对被试处理公文的方式、方法、结果等进行评价。

公文筐测验一般用于对高级管理者的评价，它可以对应试者的计划、预测、决策、沟通等管理能力进行测试，在管理领域应用十分广泛。

公文筐测验也可以算做是一种模拟情境测验，因为应试者对文件的处理结果就是待考察的工作样本。但它又和一般操作性的模拟情境测验不同。被测者并没有处在一种真实的情境中，而是通过测验指导语的说明，让被试假想自己正处于某个情境当中。

##### 二、面试

面试（interview）是考官针对自己感兴趣的、与工作有关的各种问题，与应聘者进行面对面的交谈，收集有关信息，从而达到了解、评价应聘者的目的。

面试是人事管理领域应用最普遍的一种测量形式。但严格地讲，面试并不能算是一种测量，因为面试的结果往往不能定量化，对结果的评价也不够客观。但是，因为面试可以给测试者和被测试者提供双向交流的机会，尤其能使人了解到从纸笔测验的卷面上看不到的内容，在招聘中往往被采用。面试按其形式的不同，可以分为结构化面试和非结构化面试。

##### （一）结构化面试

结构化面试（structured interview）往往有事先确定的提问提纲，里面列出需要了解的各方面的问题，而且这些问题通常还可能有一定的内在的逻辑结构。面试时，考官按照固定的程序向应试者逐个提问这些



问题。这样，所有应试者都回答同样结构的问题。结构化面试往往有标准化的评分表和详细的评分标准，考官按照评分标准对应试者的表现进行客观评分。

结构化面试克服了面谈的无法量化、主观和随意性的缺点，是一种很值得推广的面谈技术。

## （二）非结构化面试

非结构化面试（unstructured interview）则没有固定的面谈程序，面谈者提问的内容和顺序都取决于面谈者的兴趣和现场应试者的回答。这种面试方法给谈话双方以充分的自由，面试考官可以针对应试者的特点进行有区别的提问，不同应试者所回答的问题可能不同。我国很多企业的人事经理在招聘中都喜欢采用这种面试形式。

虽然，非结构化的面谈形式给面试考官以自由发挥的空间，但由于这种形式的面试存在一系列问题，如容易受考官主观因素的影响，面试结果无法量化，不利于应试者之间的相互比较等，建议尽量不要在对应聘人员的评估中采用这种方法。

在实际应用中，还存在一种介于结构化和非结构化面试两种形式之间的折中形式，即在结构化面试中加入一些灵活的追问和附加问题。这样既可以保证对应试者进行客观评估，又使考官能够根据应试者不同的情况，随时了解面试提纲之外的有用信息。这种方法称做“半结构化面试”。

## 三、调查法

在人事管理工作中，还经常要用到一种测量方法：调查法。调查法(survey)是指就员工的某些意愿、态度、观点或感受等认知性或情感性心理状态或行为倾向，进行一定范围的信息搜集，并就相应状态或倾向的特性、程度、广泛性等作出分析评价，通常由一系列问题构成，可以是纸笔形式，也可以是访谈形式。

调查表与人事测验不同。调查表没有严格的常模，也不需具备考察的信度、效度指标，它只是管理者用来了解现实情况的一种工具。例如，企业中常进行的员工满意度调查、员工价值观调查、员工需求调查、员工对某项改革措施的态度调查等，都属于此类。

调查表虽然不很严格，但它是管理工作的必备工具之一。通过对管理者自身、员工以及工作群体乃至整个组织的各方面特点和现状的评估，可以帮助管理者调整自己的管理方式，制定相应的策略，确定合适

的激励手段，并能帮助管理者及时发现组织中存在的问题，及时解决。

# 全国迷你型MBA职业经理双证班

**认证系列：**职业经理资格认证、人力资源总监、营销经理、品质经理、生产经理、物流经理、项目经理、企业培训师、酒店经理、市场总监、财务总监、营销策划师等认证。

**颁发双证：**通用高级经理资格证书+MBA 高等教育研修结业证书（含 2 年全套学籍档案）

**证书说明：**证书全国通用、国际互认、电子注册，是提干、求职、晋级、移民的有效依据

**学习期限：**3 个月（允许工作经验丰富学员提前毕业） **收费标准：**全部学费 **1280** 元

**咨询电话：**13684609885    0451- 88342620    88723232    **电子邮箱：**[xchy007@163.com](mailto:xchy007@163.com)

**学校网站：**[www.mhjy.net](http://www.mhjy.net)    **颁证单位：**中国经济管理大学    **承办单位：**美华管理人才学校

## 全国招生 函授教育 颁发双证 权威有效



职业经理 MBA 整套实战教程

千本好书 **免费** 下载    学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第四章 个性品质测验

### 第一节 卡特尔 16 种人格因素测验

#### 一、测验的功能与特点

##### （一）人格的含义

人格是稳定的、习惯化的思维方式和行为风格，它贯穿于人的整个心理过程，是人的独特性的整体写照。

##### （二）卡特尔 16 种人格因素（卡特尔 16PF）测验的功能

## 1. 维度的解释

(1) 因素 A 乐群性：表示热情对待他人的水平。

高分特征：对他人的关注程度高于平均水平，并且很容易与他人交往，对他人热情对待。

平均分特征：对他人的关注与感兴趣的程度处于平均水平。

低分特征：对工作任务、客观事物或活动所倾注的关注水平要高于对他人的关心程度。

(2) 因素 B 敏锐性：刺激寻求与表达的自发性。

高分特征：有很高的自发表达水平，思维活动非常迅速，但同时也表明，在言行之前并不总是深思熟虑。

平均分特征：表达的自然流露程度和多数人一样，在进行决策时，会进行认真思考。

低分特征：在决策之前会进行非常仔细的思考，这种深入思考的能力表明比大多数人更全面地思考，达到更深刻的理解。

(3) 因素 C 稳定性：对日常生活要求应付水平的知觉。

高分特征：感到能够控制生活的现实需要，并且能够比大多数人更沉着、冷静地应付这些要求。

平均分特征：觉得和大多数人一样能平静应付生活中的变化。

低分特征：觉得自己受到生活变化的影响很大，难以像大多数人一样沉着地应付这些生活要求。

(4) 因素 E 影响性：力图影响他人的倾向性水平。

高分特征：喜欢去影响他人。

平均分特征：并不将自己的观点、看法强加于他人。倾向于向他人表达自己的观点，但同时也让他人表达其自己的观点。当不同的观点有正确性时，愿意接受它。

低分特征：不经常表达自己对事物的看法和观点，并倾向于让他人处于领导地位。

(5) 因素 F 活泼性：寻找娱乐的倾向和表达的自发性水平。

高分特征：通常较为活泼和任性，具有高于平均水平的自发性。

平均分特征：能量水平、言行的自发性处于平均水平。

低分特征：是一个认真的，喜欢全面地思考问题的，认为别人会将其看成是一位严肃对待生活的人。

（6）因素 G 规范性：崇尚并遵从行为的社会化标准和外在强制性规则。

高分特征：崇尚社会强制性标准和规则，并愿意遵从它们。

平均分特征：倾向于接受外来强制性标准和规则，但并不僵硬地去遵从它们。有时更倾向于灵活地运用规则，而不是逐字逐句地去遵从。

低分特征：不喜欢遵从严格的规则和外在强制性指导，较之多数人更少地遵从书本原则。

（7）因素 H 交际性：在社会情境中感觉轻松的程度。

高分特征：在社会情境中比大多数人都表现自如，较之其他人更少感受到来自他人的威胁。

平均分特征：像大多数人一样，在社会情境中感到较为轻松。

低分特征：在社会情境中，尤其是在周围的人都不熟悉的情况下，会感到有些害羞和不舒服。可能自我意识较强，不喜欢被他人关注。

（8）因素 I 情感性：个体的主观情感影响对事物判断的程度。

高分特征：对事物的判断较容易受自己的情感和价值观影响。对某个决策的判断更多地基于它看起来是否正确，而不是对它进行冷静的逻辑分析。因此，在对事物进行评价时，更关注自己的品位、价值观和感觉。

平均分特征：在需要判断和决策时，倾向于注意事实以及他们的实用意义，同时也意识到有关问题的情绪性后果与价值。实际上，判断事实倾向于在主观与客观之间取平衡。

低分特征：在进行决策判断时，倾向于注重逻辑性与客观性。

（9）因素 L 怀疑性：喜欢探究他人表面言行举止之后的动机倾向。

高分特征：有一种自然倾向，认为他人的言行背后隐藏着某种动机，而不是将他人的言行按其表面意义理解。

平均分特征：倾向于认为他人是值得信任的和真诚的。可能会对值得怀疑的目的较为警觉，但当完全了解他人之后，会乐于接受他们。

低分特征：通常乐于信任他人的所说所做是真诚的，并对他人给予怀疑的信任。

（10）因素 M 想象性：个体在关注外在环境因素与关注内在思维过程两者之间寻求平衡的水平。

高分特征：勤于思考，并不拘泥事件本身的细节信息，而倾向于思索有限事实之外的东西。

平均分特征：在关注某一事件时，既关注事件的事实和细节，又会从更广阔的思路去考虑。

低分特征：是一个现实主义和脚踏实地的人，更倾向于直接去做某件事情，而不是花时间去论证其可行性。

（11）因素 N 隐秘性：将个人信息私人化的倾向。

高分特征：不愿轻易透露个人信息。似乎是一位爱保守个人秘密的人。

平均分特征：对多数人都较为公开地展示自我。

低分特征：喜欢待人公开、直率。较之大多数人来说，更乐于解释有关自己的各种信息。

（12）因素 O 自虑性：自我批判的程度。

高分特征：觉得自己有很大的困惑，或者觉得自己比别人活得更艰难。自我批判意识较强，对现实中的失误倾向于承担太多的个人责任。

平均分特征：对自己的长处和缺陷似乎有较现实的认识，能为自己的失误承担责任，能够从这些失误中吸取教训。

低分特征：和大多数人相比，很少自我怀疑。

（13）因素  $A_1$  变革性：对新观念与经验的开放性。

高分特征：对新观念与经验有强烈的兴趣。似乎对变革有很高的开放性。

平均分特征：对新观念与经验的开放程度和绝大多数人一样。

低分特征：强调按既定方法行事的重要性。和大多数人相比，很少倾向于冒险尝试新的做法与观念。

（14）因素  $E_2$  独立性：融合于周围群体及参与集体活动的倾向性。

高分特征：倾向于独立解决问题和做出自己的选择和决定。

平均分特征：力求在融合于群体及独立于群体这两个极端中寻找平衡。

低分特征：希望成为组织中的一员，并热爱组织活动。

（15）因素 Q<sub>3</sub> 自律性：认为以清晰的个人标准及良好的组织性对行为进行规划的重要性程度。

高分特征：通过对事情的事先计划和准备来对事物进行控制、有十分清晰的个人标准，并认为以此来规划自己的行为很重要。

平均分特征：对事情进行事先计划和组织的倾向同于多数人。

低分特征：不像多数人一样去对事情进行控制和进行事先的计划和组织。更乐于任由事情变化，并可以容忍某种程度上的无组织性。

（16）因素 Q<sub>4</sub> 紧张性：在和他人的交往中不稳定性、不耐烦以及由此所表现的躯体紧张水平。

高分特征：和绝大多数人相比，体验到高度的紧张，经常感受到不满和厌烦。

平均分特征：通常所体验到的躯体紧张水平和大多数人差不多。

低分特征：和大多数人相比，躯体紧张水平较低，很少感到对别人不耐烦和不满。

## 2. 测验的功能

从 16 个相对独立的人格维度对人进行描绘，可以了解应试者在环境适应、专业成就和心理健康等方面的表现。在人事管理中，16PF 能够预测应试者的工作稳定性、工作效率和压力承受能力等。可广泛应用于心理咨询、人员选拔和职业指导的各个环节，为人事决策和人事诊断提供个人心理素质的参考依据。

### （三）卡特尔 16PF 测验的特点

#### 1. 卡特尔 16PF 测验的优点

（1）高度结构化。

（2）实施简便，计分、解释都比较客观、容易。

（3）被持续证明有较好的信度和效度。

#### 2. 卡特尔 16PF 测验的缺点

（1）被试常因情境的改变而做出不同的反应，测验的信度不如智力测验等认知性测验。

（2）由于人格特质难于定义，个体行为总是受到情境与人格的交互作用的影响。

(3) 被试对问卷的回答不一定能反映其真实情况。

(4) 反应定势和反应风格影响测验结果。

(5) 相对而言，问卷的题目较多，因此被试容易出现疲劳。

## 第四章 个性品质测验

### 第一节 卡特尔 16 种人格因素测验

#### 二、测验的应用

##### (一) 适用对象

16 岁以上个体人格特征的最普遍使用的工具之一。广泛适用于各类人员，对被试的职业、级别、年龄、性别、文化等方面均无限制。

##### (二) 测验的构成

共有 16 个维度。测验由 187 道题目组成。每一个人格因素由 10—13 道测验题组成的分量表来测量，每个维度对应一个分量表，共有 16 个分量表。16 种因素的测验题采取按序轮流排列，即从第 1 题到第 16 题分别按序对应于 16 个人格因素，然后再转回来，从第 17 题到第 32 题再同样按序对应 16 个人格因素。这样既便于计分，也便于保持受试者作答时的兴趣。每一道测题有三个备选答案。

##### (三) 测验的施测过程

测验不限定时间。应试者做题时应以对问题的第一印象尽快回答，无须过多斟酌。一般用 30—45 分钟左右可以完成。

#### 1. 纸笔作答的施测过程

(1) 依据预定的参试人数选择好适宜的测验地点，布置考场。考场环境应安静整洁、无干扰、采光照明良好。

(2) 准备好测验所用的如下材料：测验题本、专用答题纸、铅笔、橡皮，保证每位应试者有以上完整的测验材料及用品。通常还有一些备用铅笔。

(3) 安排参加测验人员入场，并宣布测验注意事项。

(4) 作答时，要求应试者注意下列五点。第一，每一题目只能选择一个答案。第二，请不要费时斟酌。应当顺其自然地依您个人的反应选答。一般说来，问题都略显简短而不能包含所有有关的因素或条件。例如，有一题是有关球赛的问题，您对于观看排球赛或篮球赛的爱好可能不同，您的回答应就一般球赛而言。第三，除非在万不得已的情形下，尽量避免选择中性答案，如“介于 A 与 C 之间”，或“不确定”这样的答案。第四，请不要遗漏，务必回答所有的问题。有些问题似乎不符合于您，有些问题又似乎涉及隐私，但本测验的目的，在于研究比较青年和成人的兴趣和态度，希望应试者真实作答。第五，有些题目您可能从未思考过，或者感到不太容易回答，对于这样的题目，同样要求您做出一种倾向性的判断。

(5) 检查应试者完成了所有题目后，回收题本和答题纸，测验结束。

## 2. 计算机施测过程

(1) 启动 16PF 测验专用软件。

(2) 按照屏幕提示，输入所需要的信息，以及受试者的背景信息。

(3) 屏幕呈现答题指导，让受测者看屏幕，学会如何看题并作反应，选择符合的选项。

(4) 受试者答题。

(5) 答题结束后，出具计算机计算及报告结果，包括各个人格维度上的初步测评结果、转换后的标准分、人格因素剖面图和次元人格因素估算和应用估算分数。

## (四) 测验的样题及指导语

### 1. 指导语：

**下面是测验题。请阅读各个题目，选出符合你情况的选项。**

**每一道题后所有三个备选答案，如果你选答“A”，则在答题纸上相应的题号后涂黑“①”，选答“B”则涂黑“②”，选答“C”则涂黑“③”。**

### 2. 样题：



**(1) 我喜欢看团体球赛。**

**A. 是的    B. 偶然的    C. 不是的**

**(2) 我所喜欢的人大都是：**

**A. 拘谨缄默    B. 介于 A 与 C 之间    C. 善于交际的**

**(3) 金钱不能带来快乐。**

**A. 是的    B. 介于 A 与 C 之间    C. 不是的**

#### **(五) 结果报告**

**姓名：张君    性别：男    年龄：28    测验日期：2004 年 7 月 8 日**

**性格开朗活泼，待人热忱友善，属外向型，为人处事直爽、自然，不会刻意揣摩和计算利弊得失，对于自己无把握的事显得优柔寡断，容易随群附众，缺乏主见，但在自己所熟悉的领域则十分自信，坚持己见。决策不果断，缺乏勇气，对目标的专注程度和责任心有待提高。**

**心理健康状况良好，对竞争和压力有一定的心理适应性，能保持平和稳定的心态，在新环境中的成长能力较低，在新环境中需注意加强应变的灵活性，注意观察和分析在新旧角色之间的差别，应根据环境变化和角色要求的不同及时调整自己的情绪状态和立场观点，尽快进入新角色。**

**有一定创造能力，可以从事需要发挥个人自主性的工作，比如研究工作、设计策划等工作，事务管理能力一般，不善于处理琐碎繁杂的事务或关系，难于做有条不紊的、细致的工作，适于从事能够发挥个人才华的自主性较强的工作，不苛求于做出卓越的成就，对职业的专注程度不够，容易满足，应培养进取、负责精神，不断提高对自我的要求。**

#### **(六) 使用该测验的注意事项**

1. 使应试者真实作答。测验为自陈量表，难于控制和防止应试者的掩饰行为和倾向性作答行为，同时情境因素对应试者的影响也难以控制。为此在测验之前主试应注意向应试者做解释，尽量达成双方的信任关系，排解其顾虑和猜疑，尽可能达到一般的平和心态，同时请应试者注意在自我评述中按照自己最平常的情况作答，最大程度地反映真实状况。

2. 注意甄别应试者的反应风格，即答题倾向性，如果应试者在内容相互对抗的项目上回答相同，或

应试者在绝大多数的项目上的回答选项相同，那么有可能应试者没有认真地按自己的情况答题，是在敷衍了事或故意作假。这样的答题无效。

3. 注意考察应试者是否有社会赞许倾向，特别在用于对应聘者的筛选测验中。应聘者在高动机驱使下，会对题目有所猜测，有意或无意地改变其在测验上的反应，而塑造出一种会受到社会赞许或迎合职位需要的形象，这种形象并不能代表应试者的真实情况。对于此类人员，在结果解释中要标注出，这也代表其行为特征之一；同时应补充使用其他的评估技术，比如面试、情景模拟测验等，或在其他测验中对相应的考评要素反复考察和评定。

4. 如果需要，在本测验中可加入测谎量表，作为监控手段。实际上，16PF 测验本身就有检测被试者掩饰倾向的维度，即第 11 个因素 N，它可以反映被试是否诚实而不加掩饰地报告出自己的真实想法。

## 第四章 个性品质测验

### 第二节 DISC 个性测验

#### 一、测验的功能与特点

##### （一）DISC 个性测验的作用

DISC 个性测验着重从支配性（D）、稳定性（S）、影响性（I）和服从性（C）四个方面对人格特征进行描述，从而了解应试者的管理、领导素质以及情绪稳定性等。DISC 个性测验维度含义如下：

1. 服从性（C）：准确的、有分析力的、谨慎的、谦恭的、圆滑的、善于发现事实、高标准、成熟的、有耐心的、严谨的。

##### （1）服从性对团队的贡献

保持高标准；有责任心；善于下定义、分类、获得信息并检验信息；客观的。

##### （2）服从性的理想工作环境

需要批判性思维；技术或者专业领域；小团体的亲密关系；相似的工作环境；私人办公室或者工作环境。

(3) 服从性的压力反应

悲观的；挑剔的；紧张的；过分批评。

(4) 服从性的可能的缺陷

受到批评时采取防御措施；常常陷入细节中；对环境过分热衷；冷漠和疏远。

(5) 服从性的情绪特征：害怕。

2. 稳定性 (S)：友善的、亲切的、好的倾听者、有耐心的、放松的、热诚的、稳定的、可靠的、团队合作者、善解人意的、稳健的。

(1) 稳定性对团队的贡献

可靠的团队合作者；有耐心和同情心；逻辑性的思维；服务取向。

(2) 稳定性的理想工作环境

稳定的、可预测的环境；变化慢的环境；长期的团队合作关系；人们之间较少冲突。

(3) 稳定性的压力反应

非感情的表露者；漠不关心；犹豫不决；坚定的。

(4) 稳定性的可能缺陷

倾向于不愿意辩论；在决定优先权时遇到困难；不喜欢非正当的变化。

(5) 稳定性的情绪特征：非情绪化。

3. 支配性 (D)：爱冒险的、有竞争力的、大胆的、果断的、直接的、创新的、坚持不懈的、问题解决者、任务取向的、自我激励者。

(1) 支配性的团队贡献

基层的组织者；前瞻性的；以挑战为导向；发起运动；创新精神。

(2) 支配性的理想环境

不受控制、监督和琐碎事情困扰；革新的；有表达思想和观点的论坛或者集会；

非日常工作、挑战性和机遇性的工作。

### （3）支配性的压力反应

高要求的；紧张的；有野心的、好侵略的；自负的。

### （4）支配性的可能缺陷

过度使用地位；制定过高的标准；缺乏圆滑和变通；承担过多的工作。

### （5）高支配性的情绪特征：愤怒。

4. 影响性（I）：表现为有魅力的、自信的、有说服力的、热情的、鼓舞人心的、乐观的、令人信服的、受欢迎的、好交际的、可信赖的。

### （1）影响性对团队的贡献

乐观、热情；创造性解决问题；激励他人为组织目标奋斗；团队领导者；协商解决问题。

### （2）影响性的理想工作环境

人们之间关系密切；不受控制和琐碎事情的困扰；有活动的自由；有传播思想的论坛和集会；有相互联系的民主监督者。

### （3）影响性个体在压力下的反应

自我提高；过分乐观；过多言语；不现实的。

### （4）影响性可能的缺陷

忽略细节；在评价人方面不现实；不加区分地相信别人。

### （5）高影响性的情绪特征：乐观。

## （二）DISC 个性测验的特点

1. 适合一般、正常群体。

2. 与管理绩效相关，能很好地描述应试者的个性特征（优势、劣势等），并能预测其领导特征和情绪稳定性等。

3. 测验时间短，简单易行。

4. 有比较完善的解释体系。

## 第四章 个性品质测验

### 第二节 DISC 个性测验

#### 二、测验的应用

##### （一）适用对象

1. 大、中学生。DISC 个性测验有助于大学生和中学生科学地了解自己，从而做出正确的职业选择。
2. 社会上的一般人员。成年人往往对自己的性格特点也不是十分地了解，他们通常对个性测验比较感兴趣。通过 DISC 测验，他们可以知道自己性格上的优势、缺点，有助于更好地进行工作。
3. 管理人员。对管理人员进行 DISC 个性测验，目的就在于在工作情境中了解其优势、在工作中应注意的事项以及一些个性特征等。从而选择一些更易与企业形象、企业文化相融合的个体，形成企业独特的特征和氛围。

##### （二）测验的构成

DISC 个性测验由 24 组描述个性特质的形容词构成，每组包含四个形容词，这些形容词是根据支配性（D）、影响性（I）、稳定性（S）和服从性（C）四个测量维度以及一些干扰维度来选择的，要求被试从中选一个最适合自己的和最不适合自己的形容词。

##### （三）测验的施测过程

DISC 个性测验不限定时间，整个测验大约需要十分钟左右。纸笔作答或计算机施测均可，可以集体施测，具体过程与卡特尔的 16PF 类似，大致包括以下四个步骤：

1. 依据预定的参试人数选择好适宜的测验地点，布置考场。考场环境应安静整洁，无干扰，采光照良好。
2. 准备测验所用的如下材料：问卷、铅笔、橡皮，保证每位应试者有以上完整的测验材料及用品。
3. 安排被测评人入场，并宣布测验注意事项。
4. 检查应试者完成了所有题目后，回收问卷，测验结束。

（四）测验的指导语及样题

1. 测验指导语

**请在以下每一组词中选择一个最接近于你（最适合描述你）和一个最不接近于你（最不适合描述你）的词，用“/”标在后面相应的括号中。这不是好坏的评价，不是应该如何，而是对自己的客观描述。24组词是相互独立的，每一组都要回答，不要有遗漏。请请合作。**

2. 测验的样题

	最接近	最不接近
温柔	[    ]	[    ]
说服力	[    ]	[    ]
谦卑	[    ]	[    ]
独创性	[    ]	[    ]

（五）测验报告的样例

**被测评人编号：001**

**姓名：\*\*\***

**性别：男**

**年龄：33**

**百分等级：支配性 D=42.2，影响性 I=73.6，稳定性 S=58.3，服从性 C=77.6。**

### **1. 优势**

**注重质量，是高质量的促进者和支持者，**

**对保持高标准能很好地把握分寸，有轻重缓急的判断能力，**

**好组织、好管理，欣赏团队中有相同质量意识的个体，**

**对社会和工作环境的变化比较敏感。**

### **2. 需要注意的事项**

**应更多地接受他人的思想和观念，**

**要设置比较现实的目标，**

**不要过分计较他人的评论。**

### **3. 个人倾向**

**目标：乐观的，**

**通过什么标准判断他人：自尊和成就，**

**通过什么影响他人：好的人际关系中的策略，**

**对团队的贡献：创造好的工作环境，**

**经常并过度使用：机智和圆滑，**

**处于应激状态时：太温和，讨好人，**

**害怕：为了取得好的关系而放弃质量。**

## 第四章 个性品质测验

### 第三节 管理人员人格测验

#### 一、测验的功能与特点

##### （一）管理人员人格测验的功能

## 1. 基本功能

该测验从正性情绪倾向、负性情绪倾向、乐群性、责任心、广纳性、内控性、自控性、自信心、A型人格、成就动机、权力动机、面子倾向这12个与管理绩效有关的人格特点对人进行描绘，了解应试者在行为风格、思维和处事方式方面的特点，并且能够预测应试者的组织管理潜力，对其适宜的发展方向和组织环境提出建议。该测验适合对管理人员的选拔和职业指导，为人事决策和人事诊断提供参考依据。

## 2. 维度含义的解释

(1) 正性情绪倾向：倾向于体验到正性情绪，对自己感觉良好，也称为正性影响力。

正性情绪得分高的个体表现为：

较社会化、亲切、友好；

在工作上，更容易体验到正性情绪，对工作感到满意；

更乐于和同事交往，通常在组织中和其周围的人群相处很好；

通常在需要高社会交往的工作岗位上绩效较好，例如，销售部门和客户联络部门。

低分个体：

较少体验到正性情绪；

与他人的社会交往较少。

(2) 负性情绪倾向：倾向于体验到负性情绪，用负性的眼光看待自我和周围世界的人格特质，也称为负性影响力。

高分个体：

更倾向于体验负性情绪和感觉到时间、环境的压力；

更倾向于在工作上体验到负性情绪，感到压力；

对自己和绩效要求更为严格，这种倾向使他们趋向于去提高自己的绩效；

往往通过唱反调和指出议题的反面效果来使团体决策更为谨慎自制；

他们可能在那些需要批判性思考和评估的位置，例如质量监测部门，有较好的绩效。



低分个体：

较少感受到环境中的压力；

有较强的耐受力，能承担风险较大、容易受挫的工作；

有时可能会过度乐观。

（3）广纳性：个体有独创性，广泛接受各种刺激，有广泛兴趣，愿意冒险。它的对立面是思维狭隘，小心谨慎。在组织中广纳性往往被解释为有创造力和革新性的行为。

高分个体：

对变化性大、需要创新性或较为冒险的工作较为适宜；

企业家往往具有高广纳性。

低分个体：

相对较为保守和谨慎；

倾向于依赖过去的经验和规则，不愿积极创新，不希望承担改革的风险。

（4）责任心：个体表现为认真、审慎和坚忍的倾向。在许多组织环境中，责任心都是十分重要的。

高分个体：

有组织性和自律性；

在大多数组织的大多数工作中，责任心都可预示较高的绩效。

低分个体：

可能缺乏方向性和自律性；

可能耐心不足，计划性和条理性较差；

倾向于采取较为灵活的办事方法。

（5）乐群性：与他人相处融洽的倾向。乐群性是一个用来区分个体是否善于与人相处的人格特质。

高分个体：

通常受人喜爱，善于照顾他人，对他人亲善；

通常易于相处，是好的团队合作者；

在需要与他人发展良好关系的岗位上，其特性是一项宝贵资源。

低分个体：

往往不招人喜爱，对人不信任，没有同情心，不合作并且粗鲁；

或许在那些需要体现对抗性的职位上，例如，作为追债员或训练员，会有一些优势。

（6）内控性：描述人们相信命运、运气或一些外部力量对他们周围所发生的事情起决定性作用的维度，反映了人们关于自己对周围环境或身边所发生的事情到底有多大控制力的看法。

高分个体：

相信他们能影响周围的世界并控制自己的生活；

认为他们自己的行为会影响自己的经历；

如果他们取得好的成绩，趋向于将成绩的取得归结于自己的素质，例如他们的能力或努力；

更容易被激励，而且不需要过多的直接指导，因为他们更相信他们的工作努力会带来好的结果，例如，他们努力工作会使工资增加，受到奖励，获得工作安全感，得到提升。

低分个体：

认为他们对环境的影响很小，并对自己所经历的大事也没有什么控制力；

倾向于相信外部力量决定他们的命运；

他们认为自己的努力与自己将要经历的事情的关系不大；

倾向于将自己的成功归结于外部力量，例如运气、有权势的人的作用，或仅仅是因为任务简单而已，他们会将原因归结于组织。

（7）自控性：人们试图控制自己在他人面前的行为方式的倾向。

高分个体：

希望他们的行为会为生活所接受，并且他们会调整自己的行为以求和社会赞许线索相适应；

力求根据所处的环境表现出恰当的行为，更倾向于根据所处的环境调整自己的行为；

很善于处理他人对自己的印象；

在那些需要与各种人打交道的职位（例如，销售或咨询）上，可能会有更好的绩效；

因为他们会调整自己的行为以适合所处的群体，当组织需要有人与那些对组织提供支持的外部团体交涉时，他们往往能担当重任。

低分个体：

对暗示可接受行为的线索并不敏感；

对表现与情境相符合的行为并不在意，例如，他们可能在一个有领导参加的会议上表现出不耐烦，或告诉顾客对自己有竞争力的商品的优点；

往往由自己的态度、信念、情感和原则所引导，并不关心他人对其行为的看法；

更倾向于认为他们所想的是真实的或正确的，并且不太关心其他人对他们的反应；

可能在需要组织成员给予开诚布公、诚实的反馈的职位上更为适合（尤其是当反馈是负面的时候），并且他们可以在决策群体中扮演唱反调的角色。

（8）自信心：人们为自己和自己的能力感到自豪的倾向。了解自信心有助于在管理中对行为的理解。自信心会影响人们对活动和工作的选择。

高分个体：

认为他们通常较能干，是能应付大多数情境的有价值的人；

更倾向于选择有挑战性的工作和职业；

在工作时会为他们设置更高的目标，并且更喜欢去处理那些较困难的工作；

对工作动机和工作满意度有正面的影响。

低分个体：

往往对他们的自我价值提出质疑，怀疑自己的能力，担心自己是否能在不同的努力中取得成功；

不论他们如何怀疑他们的努力，他们的能力与高自信性的个体没有什么区别。

（9）A 型人格：

## A 型人格

有强烈的成就动机和竞争意识，并有强烈的紧迫感；

往往有强烈的成就需求、竞争意识、紧迫感，较为急躁和有敌意；

有一种要在较短时间做大量工作的要求；

往往较难相处，A 型管理者比 B 型管理者更容易与其属下和同事发生冲突；

不适合做那些需要大量人际间交流的工作；

他们常常因为不耐烦而打断别人并帮他们续上后半句话；

不是好的团队合作者，更适合于单独工作；

他们急于看到结果，容易在长时段的项目中受到挫折。

## B 型人格

倾向于温和、宽容、放松；

倾向于采取慢节奏、较为轻松随意的工作方式；

适合于随意性较大、压力较小的工作。

（10）成就动机：喜欢接受挑战性的任务，希望达到个人的高目标的倾向，企业家和管理者通常有较高的成就动机。

高分个体：

希望接受挑战性任务，并且喜欢将个人目标定得较高；

喜欢对所发生的事情负责，例如，为他们自己设定明确的目标，并愿意为事情的后果负责，而且他们希望得到反馈；

他们通常倾向于从事能满足他们强烈的成就需要的工作；

常有很强的目标方向性，并趋向于选择中等水平的风险。

低分个体：

追求个人高目标的愿望不强；

倾向于容忍自己的失败；

喜欢较为轻松、没有压力的生活。

（11）权力动机：希望控制或影响他人的行为、情绪的倾向。

高分个体：

有想对他人进行情绪、行为上的控制和影响的强烈愿望；

倾向于处在那些需要表现对他人的影响力的职位上，例如，管理职务和领导位置；

作为领导通常比那些低权力动机的个体做得更好；

通常不宜让两名有强烈权力动机的个体处在同一情境中，这样他们都希望体现其影响力，容易导致冲突。

低分个体：

对他人的依赖性较强，希望在他人的指导下工作；

对他人的工作有较强的认同感，是一名较好的合作者；

满足于自己的现状，不愿和他人产生权力之争。

（12）面子倾向：看重面子问题，一方面维护自己的面子；另一方面为顾及情面为他人留面子，不做自认为使自己或他人丢面子的事情。

高分个体：

力求受到他人的重视、赞赏、推崇，希望能在别人心目中占重要地位及留下美好印象，从而在内心产生荣耀、光彩、神气、得意等愉快感觉；

当他们感觉到不受他人尊重、在竞争中失败、暴露自己的缺点、失态、不受欢迎等，会在内心中产生难堪、困窘、尴尬、羞耻等不快感觉；

为了防止丢面子或为了追求有面子的事，会表现出一些并非表里如一、名副其实的言行，目的是操纵留给他人的印象。

低分个体：

对人际关系、他人感受的敏感性较低；

相对较为坦率，但可能会由于直接表达自己的观点而伤害他人；

不在意他人对自己的评价，倾向于表现其真实感受。

## （二）测验的特点

该人格测验是根据“大五”人格模型中五种人格维度即正性情绪倾向、负性情绪倾向、乐群性、责任心、广纳性及经大量研究证明与管理绩效有关的 7 项人格维度（内控性、自控性、自信心、A 型人格、成就动机、权力动机、面子倾向）构建成的一项个性测验。

1. 测验的优点：高度结构化，实施简便，计分、解释都比较客观、容易。

2. 测验的缺点：第一，被试常因情境的改变而做出不同的反应，测验的信度不如智力。

测验等认知性测验。第二，由于人格特质难于定义，个体行为总是受到情境与人格的交互作用影响。

第三，被试对问卷的回答不一定能反映其真实情况。第四，反应定势和反应风格影响测验结果。

## 第四章 个性品质测验

### 第三节 管理人员人格测验

#### 二、测验的应用

##### （一）适用对象

本测验适用于对管理人员的评估，以及对管理岗位候选人员的甄选。该测验广泛适用于各种岗位，对对象的职业、年龄、性别、文化等方面均无限制。

##### （二）测验的构成

该测验针对 12 个人格维度进行考察。测验题目以三择一的选择题形式出现。测验共有 145 题，每一人格因素由 10~15 个题目组成的量表来测量，12 种因素的测试题采取按序轮流排列，以便于计分，并保持受试者作答时的兴趣。

##### （三）测验的施测过程

与前面两个测验是类似的。

#### （四）测验的指导语和样题

##### 1. 测验的指导语

**本测验包括一些有关个人态度与态度等方面的问题。每个人对这些问题的看法不同，回答也就自然会有所不同。因此对问题如何回答，并没有“对”与“不对”之分，只是表明你对这些问题的态度。请您尽量表达您个人的真实意见。**

**每一问题都有三种可选择的回答。请您从中选出最适合于您的答案，并在答题纸上把题号后把与您的回答相对应的数字圆圈涂黑。**

**例如，若您对某题选择答案 1，则在答题纸上把该题号后的①涂黑。**

**请您根据自己的实际情况回答。对每个问题不要过多考虑，请尽快回答。**

**每个问题都要回答，不要有遗漏。对每个问题只能选择一种回答。**

##### 2. 样题：

**（1）在开会时，如果我的意见与领导不同，我往往，**

- A. 保留自己的意见，一般不与之争论**
- B. 依情况而定**
- C. 当场表明立场**

**（2）如果让我组织一个社团活动，我认为，**

- A. 需要有经验的人给予很多协助**
- B. 自己有能力办得有声有色**
- C. 依情况而定**

#### （五）报告样例

该应试者正性情偏明显，对生活乐观，对竞争和压力有一定的心理适应性，能保持平和稳定的心态，自信心较强，相信自己的能力，自控性较强，能较好地处理自己在他人面前的印象，有很高的成就动机，希望能做出卓越成就，对职业的专注积极性较高，积极进取，不断提高对自己的要求，有很强的责任心，办事认真、审慎，对有组织性和原则性的事务管理能力较强，善于处理琐碎繁杂的事务或关系，能够做到有条不紊的、细致的安排，适于从事需要耐心和审慎的工作，广纳性较强，有较强的开创精神，容易接受新事物，有一定创造能力，可以从事需要发挥个人自主性的工作，比如研究工作、设计策划、编辑等工作。

#### （六）测验的注意事项

1. 使应试者真实作答。
2. 注意甄别应试者的反应风险，即答题倾向性。
3. 注意考察应试者是否有社会赞许倾向，特别在用于对应聘者的筛选测验中。
4. 如果需要，在本测验中可加入测谎量表，作为监控手段。



职业经理 MBA 整套实战教程

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第五章 职业适应性测验

### 第一节 生活特性问卷

#### 一、测验的功能与特点

##### （一）什么是动机

动机，是行为的内在原因，主要指发动一定的行为满足某种需要的意愿，它由需求而产生，为行为提供能量，具有目标指向性。

##### （二）生活特性问卷的功能

##### 1. 测验的功能



生活特性问卷是为评定个体的动机水平而编制的。测验从权力动机、成就动机、亲和动机、风险动机四个方面描述应试者的动机模式和强弱程度。

在个体层次上，该测验有助于预测员工的行为表现、稳定性，为有效地控制管理人员和选拔合格应聘者提供重要的信息。在团体层次上，使用动机测量工具可帮助管理者了解、控制组织成员的动机结构、水平，并进行有效的激励，以提高组织绩效和员工满意度。

## 2. 维度的解释

（1）风险动机：指决策时敢于冒险，敢于使用新思路、新方法；不惧怕失败的动机。

高风险动机的人可能过于莽撞，对可能的危险和损害估计不足，缺乏足够的大局意识和责任感，缺乏对失败的应变策略；低风险意识的人则过于保守、审慎，优柔寡断，谨小慎微，缺乏决断。

（2）权力动机：指人们力图获得、巩固和运用权力的一种内在需要，是一种试图控制、指挥、利用他人行为，想成为组织的领导的动机。

高权力动机的人往往有许多积极有利的特征，例如，善于左右形势大局，果断自信，试图说服人；但权利动机过高的人也可能会成为组织中的危险人物，他们可能只顾及个人权力，在极端情况下会不择手段，不顾组织的利益，甚至危害组织。总的来说，权力动机是有价值的，一定水平的权力动机是企业管理者实现统率力的行为根源，同时在组织中要控制权力动机的无限扩张。

（3）亲和动机：指人对于建立、维护、发展或恢复与他人或群体的积极情感关系的愿望。其结果是引导人们相互和睦、关心，形成良好的人际氛围。

亲和动机强的人能很容易地与他人沟通、交流，并且促进团队中积极的社会交往；他们富有同情心，容易接纳他人，减少冲突，避免竞争，有利于合作气氛。亲和型的领导受下属的接受和拥护，团队合作密切。但亲和动机过于强烈时可能有副作用，如回避矛盾、害怕被拒绝；过于求同，忽视个性，甚至息事宁人，放弃原则。

（4）成就动机：指人们发挥能力获取成功的内在需要，一种克服障碍，完成艰巨任务，达到较高目标的需要。是对成功的渴望，它意味着人们希望从事有意义的活动，并在活动中获得完满的结果。

由于成就动机具有行为驱动作用，在智力水平和其他条件相当的情况下，高成就动机的人获得成功更大、绩效更突出。但成就动机过高也有逆反现象：人们对目标的设置降低难度，倾向于回避失败，结果是动机的行为驱动力减退，工作任务未必尽善尽美。而且，害怕失败就害怕尝试多种可能性，无形中放弃、丧失很多机会。

### （三）生活特性问卷的特点———坚实的理论基础

本测验从近代激励理论关于员工行为动机的基本概念出发，以风险动机、权力动机、亲和动机和成就动机为维度构建而成。这些维度与人们的工作绩效以及参与管理活动中的效能是有必然联系的。

成就动机和工作绩效之间有高相关，高成就动机有利于实现高水平的个人绩效，但不一定是出色的经理人；而亲和动机和权力动机与管理绩效有密切的关系，高权力动机是管理效能的一项必要条件，同时要求亲和动机较低。

## 第五章 职业适应性测验

### 第一节 生活特性问卷

#### 二、测验的应用

##### （一）适用对象

本测验广泛适用于各行业、各层次人员，特别是面临择业、改行或求职的应聘者，用于评估其动机与职业的匹配程度。

##### （二）测验的构成

本问卷测试的是风险动机、权力动机、亲和动机、成就动机四种动机。每种动机选定 11—15 道题目加以测试。

每道题目陈述一个观点，应试者根据他对此观点的同意程度用七点量表（相当于 7 分制）评分，如“完全同意”评“7”分，“完全不同意”评“1”分。将题目随机排列编成生活特性问卷，问卷由 51 道题目组成。

##### （三）测验的实施过程

测验不限定时间，要求应试者凭直觉作答，不用过多考虑，一般在 20 分钟左右可以完成该测验。测验的实施过程与以前讲过的个体测验非常相似。

#### （四）测验的指导语及样题

##### 1. 测验的指导语

这是一份关于个人观点的问卷调查，请根据你对每一个问题的看法用 1-7 的分值表示你的同意或不同意的程度（1 代表“完全不同意”，7 代表“完全同意”）。请将你的回答填在相应问题前的横线上。答案无对错之分，请不要顾虑。每个问题都要回答，不要遗漏。谢谢合作。

##### 2. 测验的样题

1	2	3	4	5	6	7
完全	非常	稍有	无所谓	稍有	比较	完全
赞同	赞同	赞同		赞同	赞同	赞同
_____我喜欢对他人的工作做指导。						

#### （五）测验报告举例

编号: 1      姓名: 111      性别: 男      年龄: 35

#### **1. 风险动机**

**风险动机指决策时敢于冒险, 敢于使用新思路、新方法, 不惧怕失败的动机。**

**应试者的得分属中等水平, 说明一定程度上肯于创新, 做事大体上比较果断, 能冒一定程度的风险, 对可能的失败有所准备。适合担任组织的决策者。**

#### **2. 权力动机**

**权力动机指试图控制、指挥、利用他人, 想成为组织的领导的动机。**

**应试者的得分属于中等水平, 说明在一定程度上试图去影响他人, 有时喜欢领导、指挥他人, 出谋划策, 竞争意识中等, 比较自信, 一定条件下能承担责任, 适合担任组织的领导。**

#### **3. 亲和动机**

**亲和动机指对于建立、维护、发展或恢复与他人或群体的积极情感关系的渴望。**

**应试者的得分较高, 说明看重友情, 易于和他人沟通、交流、有同情心, 容易接纳他人, 有利于团队的合作气氛, 但可能回避矛盾, 过于求同, 甚至息事宁人, 应注意不要放弃原则, 才能做好评价他人的工作或监督工作, 如果是管理者, 常受到下属的接受和拥护。适合担当团队的组织者和社交性职务。**

#### **4. 成就动机**

**成就动机指个体发挥能力、克服障碍, 完成艰巨任务, 获得一定成就的需要。**

**应试者的得分较高, 说明喜欢面对困难和考验, 认为难于完成的任务更能培养自己的能力, 更能体现自己的价值。非常想发挥自己的潜力, 做出卓越的成就。适合当经理人和企业家。**

### **(六) 测验需要特别注意的事项**

1. 这里用“生活特性”代替“动机”, 是因为动机一词既敏感, 又不易被一般人理解, 可能造成应试者较多的猜测。
2. 不同职位(岗位)要求从业者的适宜的动机状态各有不同, 表现在各种动机的重要程度不同, 每一项动机的适宜定位水平不同, 因此, 不能统一地设定动机水平优劣的分数标准。
3. 针对不同类型的工作对素质或素质组合及水平的需要, 考察具体职位的动机要求, 确定考察动机

水平的标准与具体职位的对应关系。一般而言，高层管理者的权力动机要比较高、成就动机中等适度偏高，亲和动机中等；中层管理者的成就动机应比较高，权力动机应中等适度偏高，亲和动机中等；职位越低的管理者，权力动机水平应当逐渐降低，亲和动机水平应逐渐升高。

## 第五章 职业适应性测验

### 第二节 需求测试

#### 一、测验的功能与特点

##### （一）什么是需求

需求是人的个体的和社会的客观需求在人脑中的反映是缺失并追求满足的状态。

需求是动机的基本来源；动机的产生原因就是需求的满足。需求是决定行为目标的根本原因。

##### （二）测验的功能

需求测试是测查应试者对生理需要、安全需要、归属和爱的需要、自尊的需要和自我实现的需要等各大类生活需要的程度，可全面了解个体的需求状况和需求的主次形态，并可定性、定量分析员工总体需求分布模式以及各种需求的强弱程度。

在团体层次上，通过对组织全体员工实施需求测试，可揭示各层次员工的需求结构，根据这个结构可了解团体中需要的分布、形态，这是安排组织激励、调度员工士气的基本环节。需求测试和价值取向评估相互对照使用，可为组织人事工作、动机激励、企业文化建设提供依据。

五种需要的定义如下：

1. 生理需要：指各种用于满足生存的基本物质需要，如饮食、睡眠、营养等。
2. 安全需要：指对安全、稳定、依赖的需要，希望免受惊吓、焦躁和混乱的折磨，有稳定的工作等。
3. 归属和爱的需要：指对爱、情感、友谊、归属和社会交往的需要。希望拥有朋友、爱人和亲人。

如果得不到满足，个体会感到孤独。

4. 自尊的需要：指对于自己的稳定的、牢固不变的、较高的评价的需要或欲望，对于自尊、自重和

来自他人的尊重的需要或欲望。

5. 自我实现的需要：指个体充分发挥自己的潜能，实现人生价值的需要，也就是说一个人生下来具有什么样的潜能，他就希望成为什么样的人。

### （三）测验的特点——以马斯洛的需求层次理论为基础

本测验的设计和建构参照了马斯洛的需求层次理论所提出的五种层次的需求形式，以生理需求、安全需求、爱的需求、自尊的需求、自我实现的需求为维度建构而成。其中生理、安全需要、爱与归属的需要为较低层次的需求，主要靠外在的事物来满足，在马斯洛的理论框架中称为缺失需要；而自尊的需要、自我实现的需要为高层次需求，是自己内在的满足，在马斯洛的理论框架中称为生长的需要。

## 第五章 职业适应性测验

### 第二节 需求测试

#### 二、测验的应用

##### （一）适用对象

本测验广泛适用于任何希望了解自我状态的个体和各行业、各层次人员。同时，适用于对组织全体在职人员的集体施测，可了解各级员工的需求构成，为实施有效激励措施提供建议和依据。

##### （二）测验的构成

测验根据马斯洛的需求层次理论编制，以该理论体系中的五种基本需求：生理需求、安全需求、归属和爱的需求、自尊的需求、自我实现的需求作为测验维度。每种需求选定 10—16 道题目加以测试。

每道题目陈述一个观点，应试者根据他对此观点的同意程度分七点评分，如“完全同意”评“7”分，“完全不同意”评“1”分。将题目随机排列编成需求测试问卷，测验由 67 道题目组成。

##### （三）测验的施测过程

测验不限定时间，要求应试者凭直觉作答，不用过多考虑，测验时间约为 30 分钟。纸笔或计算机作答，可集体施测。具体的实施过程与以前所讲的测验相同。

#### （四）测验的指导语以及样题

##### 1. 测验的指导语

这是一份关于个人观点的测验调查，请根据你对每一个问题的看法用 1—7 的分值表示你的同意或不同意的程度（1 代表“完全不同意”，7 代表“完全同意”）。请将你的回答填在相应问题前的横线上。答案无对错之分，请不要顾虑。每个问题都要回答，不要遗漏。谢谢合作。

##### 2. 测验的样题

1—— 2—— 3—— 4—— 5—— 6—— 7

完全不 非常不 稍有不 无所谓 稍有 比较 完全  
赞同 赞同 赞同 赞同 赞同 赞同 赞同

\_\_\_\_\_ 1. 我认为有一个安稳的住所是发展事业的前提。

#### （五）报告样例

姓名：\*×\* 性别：女 年龄：25

**生理需要：53.5**——应试者得分属中等水平，说明希望得到适当的物质待遇和工作条件，在择业中有所考虑但并不十分苛求，只要达到一般水平即可。

**安全需要：88.3**——应试者得分属较高的水平，说明认为生活中会有种种危险和潜在危险，担心自己的人生安全、身体健康，害怕失业，希望有稳定的工作，安全的生活工作环境。

**归属与爱的需要：74.8**——应试者得分属中等水平，说明对友情和他人的关心重视程度一般，虽希望融入一个团体，希望与同事和上级有良好的人际关系，但并不十分考虑。

**自尊的需要：45.7**——应试者得分偏低，说明不特别在意他人对自己的评价，有自己相对独立的观点和方式，也不太在乎自己的身份地位究竟如何。

**自我实现的需要：55.4**——应试者得分属中等水平，说明对发挥自己的才干、提高自己的能力的愿望并不十分迫切。

#### （六）测验需要特别注意事项

该测验的题目设计是假定受测者愿意并能够直率地表达自己内心的真正想法，并不排除受测者出于某

些复杂的原因掩饰自己真实的想法。因此，本测验适用于企业的激励设计，员工民意调查，而不太适合于选拔。

## 第五章 职业适应性测验

### 第三节 职业兴趣测验

#### 一、测验的功能与特点

##### （一）什么是兴趣

兴趣是重要的心理特征之一，是个体力求认识某种事物或从事某种活动的心理倾向，表现为个体对某种事物、某项活动的选择性态度或积极的情绪反应。

##### （二）测验的功能

职业兴趣反映了职业（工作活动）特点和个体特点之间的匹配关系，是人们职业选择的重要依据和指南，职业兴趣测验正是用于了解这两方面特点之间的匹配关系，从而为实现“恰当的人从事恰当的工作”提供可靠的科学依据。

从个人择业方面来说，职业兴趣测验是帮助人们明确自己的主观性向，从而能得到最适宜的活动情境并给予最大的能力投入。从管理的角度来看，职业兴趣测验在管理人员的选拔和安置方面也起着举足轻重的作用。同时，职业兴趣测验还可以在能力鉴定的基础上甄别可能取得最大效益和成功的活动（职业）。该测验从以下几个方面来考察人的兴趣：

1. 艺术取向：喜欢艺术性工作，这种取向类型的人往往具有某些艺术技能，喜欢创造性的工作，富于想象力。这类人通常喜欢同观念而不是事务打交道。他们较开放、好想象、独立、有创造性。
2. 事务取向：喜欢传统性的工作，这种人有很好的数字和计算能力，喜欢室内工作，乐于整理、安排事务。
3. 经营取向：喜欢诸如推销、服务、管理类型的工作，这类人通常具有领导才能和口才，对金钱和权力感兴趣，喜欢影响、控制别人。



4. 研究取向：喜欢各种研究性工作，这类人通常具有较高的数学和科学研究能力，喜欢独立工作，喜欢解决问题；喜欢同观念而不是同人或事务打交道。

5. 技能取向：喜欢现实性的实在的工作，这类人通常具有机械技能和体力，喜欢户外工作，乐于使用各种工具和机器设备。这种人喜欢同事务而不是同人打交道。他们真诚、谦逊、敏感、务实、朴素、节俭、腼腆。

6. 社交取向：喜欢社会交往性工作，这类人通常喜欢周围有别人存在，对别人的事很有兴趣，乐于帮助别人解决难题。这种人喜欢与人而不是与事务打交道。他们往往助人为乐、有责任心、热情、善于合作、富于理想、友好、善良、慷慨、耐心。

### （三）测验的特点

1. 汲取了国外以往测验的优点和其职业兴趣的理论框架，根据中国人以及中国职业的特点设计，符合中国的国情。

2. 题目的设计具有隐蔽性，尽量避免使受测者有对职业的优劣感，降低了社会赞许性。

3. 各维度的题目合并混排，降低了测验题目维度属性的可猜测性，适合各类人员测评的需要。

4. 针对管理人员而言，可区分管理人员与非管理人员的兴趣模式，特别是不同类型管理工作所需要的职业兴趣模式。

## 第五章 职业适应性测验

### 第三节 职业兴趣测验

#### 二、测验的应用

##### （一）适用对象

职业兴趣测验适用范围很广，包括大、中学生，社会上的一般人员和管理人员。

##### （二）测验的构成

职业兴趣测验根据中国职业自身的特点，在吸收国外测验的优点和其职业理论的基础之上，采取艺术

取向、事务取向、经营取向、研究取向、操作取向、社交取向六种职业偏好作为测量维度。六种兴趣维度共设计 60 个题目。其中包括活动、工作、技能三个方面。题目的设计尽量具有隐蔽性，尽量避免使受测者有对职业的优劣感。同时各维度的题目合并混排，降低测验题目维度属性的可猜测性，以适合于管理人员测评的需要。

（三）测验的施测过程

职业兴趣测验不限定时间，六个维度 60 个题目大约 10 分钟左右的时间。测验有两种实施方式：纸笔作答或计算机施测。具体的实施过程同前面所讲的问卷。

（四）测验的指导语与样题

1. 测验的指导语

**本测验用于帮助了解你对职业的偏好，从而提供求职的咨询。每一个题目都给出一种活动或一种技能或一种职业，请用 5 分制描述你是否喜欢该项活动，或是否擅长或希望学习这种技能，或是否乐于选择这种职业。请将你的答案填在相应问题前的横线上。**

**你的回答没有正误之分，请不要顾虑。每个题目都要回答，不要遗漏。谢谢合作。**

**1————2————3————4————5**

**非常不喜欢      稍有不喜欢      无所谓      稍有喜欢      非常喜欢**

2. 测验的样题

_____ 经营取向	策划企业活动
_____ 社交取向	参加联谊活动
_____ 艺术取向	学习美术
_____ 研究取向	从事科学研究活动
_____ 技能取向	家用电器维修
_____ 事务取向	登记考勤

（五）报告样例

编号: 1 姓名: "" 性别: 男 年龄: 33

经营型 69, 社交型 65, 艺术型 46, 研究型 68, 技能型 61, 事务型 38

测验结果表明, 此人对很多方面的职业都有几乎相同程度的兴趣水平, 呈现天花板似的结果。也就是说, 此人的兴趣分布比较广泛, 同时对许多方面都很投入, 这样就可能会分散该被测试者的心理资源, 如果调节不当, 有可能会妨碍他在这些方面的潜在成就。

造成这样的结果, 可能的确是出于心理能量丰富, 资源充沛, 精力旺盛, 或者也可能出于资质聪颖, 对许多不同方面的学习、技能、活动都能驾轻就熟而并不感到任何困难, 或者也可能是出于热情, 缺乏理性的思考和抉择。一般来说, 最理想的还是将对不同方面的兴趣区分出不同水平的层次, 形成有主有次、有重有轻的格局, 这样能更好地、更经济而科学地分配有限的心理资源, 从而确保在工作、生活最重要的方面的心理投入。



职业经理 MBA 整套实战教程

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第六章 能力测验

### 第一节 多项能力与职业意向咨询

#### 一、测验的功能和特点

##### (一) 目的

多项能力与职业意向咨询从与职业活动有关的六个维度对人进行测评, 能显示出个体的相对能力强弱, 并给出个体适宜的职业排序, 从而为职业咨询提供信息。

##### (二) 功能

1. 多项能力与职业意向咨询选取了和社会的大多数职业活动有着密切关系的六个维度进行测评, 即语言理解和组织能力、概念类比能力、数学能力、抽象推理能力、空间推理能力、机械推理能力。

2. 结果显示为个体在各项能力的相对强弱、并给出适宜的职业排序，从而为职业咨询、分类和人员安置提供了科学可靠的信息。

### （三）特点

能力倾向测验（Aptitude test）一般指测量从事某种职业或活动的潜在能力的评估工具。能力倾向不是特殊训练的结果，它存在于某种特定的训练与学习之前。可以把能力倾向测验分为三类：一般能力倾向测验、特殊能力倾向测验和多因素能力倾向测验。其中一般能力倾向测验现在就是指智力测验。而狭义的能力倾向测验专指后两者。

本测验就是一种多因素能力倾向测验，它建立在可靠的心理学原理基础之上。从心理学角度说，一个人的职业选择应当是以基本能力（能力的结构与水平）及职业兴趣和性格为依据的。兴趣决定了择业的方向以及人在该方向上乐于付出的努力程度；能力，尤其是能力的不同方面的水平，则决定了一个人是否符合特定职业的要求以及是否能取得成功的可能性。

可以从许多不同的角度对能力进行分类。比如，按一般性和特殊性，可以分为一般能力和特殊能力。一般能力指通常所说的记忆力、注意力、想象力等；也就是人们常说的“智力”。特殊能力指音乐能力、绘画能力等。按能力的时间表现，可以分为“造诣”和“潜能”。造诣是人已经取得的能力成就。潜能则是指人们尚未发展、发挥出来的能力。

本测验具有以下特点：

1. 测验的结果不是一个总分或智商，而是一组不同能力倾向的分数，它提供了一种智能剖面图，显示了个体在以上六项能力的强弱分布。
2. 根据剖面图上的强弱分布，给出适宜的职业排序，并指出最适宜的职业（即能力足以胜任并使能力得以充分发挥的职业）应具备的教育水平和关键能力，从而为职业咨询提供依据。
3. 可以在计算机上进行，也可以纸笔方式作答，能很方便地进行团体施测。

### （四）维度定义

多项能力与职业意向咨询选取了和社会的大多数职业活动有着密切关系的六个维度进行测评，即语言

理解和组织能力、概念类比能力、数学能力、抽象推理能力、空间推理能力、机械推理能力。具体定义如下：

1. 语言理解和组织能力：考察对语言表达的基本理解，对语法规则、语义、语言习惯的熟练掌握程度。
2. 概念类比能力：考察对概念关系的理解，对逻辑的理解和进行类比的能力。
3. 数学能力：考察对数字关系的理解和掌握，对各种运算规则的熟练运用和对各种数学现象的敏感能力。
4. 抽象推理能力：考察对事物变换所反映出的内在规律的敏感性和对事物的抽象、概括的逻辑分析能力。
5. 空间推理能力：考察对图形进行表象加工、旋转的能力，尤其是考察人们通常所指的空间认知和形象思维的能力。
6. 机械推理能力：考察人们对一般自然常识、物理现象的认识水平，考察人对基本的物理规律和机械规则的敏感性和掌握程度。

## 第六章 能力测验

### 第一节 多项能力与职业意向咨询

#### 二、测验的应用

##### （一）适用对象

多项能力与职业意向咨询的适用对象包括：

1. 大、中学生。针对大、中学生而言，多项能力与职业意向咨询可以帮助他们了解自己各项能力的优劣，从而为升学、报考志愿提供咨询和帮助。
2. 社会上的一般人员。对于那些正处于最初择业阶段的人，即使是成年人，他也不一定了解自己胜任哪些职业，因此多项能力与职业意向咨询可为他们正确地选择职业提供科学可靠的测评数据。

3. 特别需要注意：本测验不适用于中高层管理人员。对于中高层管理者，他们的职位要求的是另一类能力，而这里所测验的基本的能力倾向和他们的管理绩效并没有太大的关系。如果拿这套测验去测查中高层管理者，很有可能他们的分数会很低，但这绝不意味着他们能力差，而是他们已不适应这一类测验题型。

## （二）测验的构成

多项能力与职业意向咨询选取了和社会的大多数职业活动有着密切关系的六个维度进行测评，即语言理解和组织能力、概念类比能力、数学能力、抽象推理能力、空间推理能力、机械推理能力。其中语言理解和组织能力有 20 道题目，概念类比有 50 道题目，数学能力有 40 道题目，抽象推理有 45 道题目，空间推理有 60 道题目，机械推理有 70 道题目。

## （三）测验的施测过程

本测验共有六个分测验，每一部分测验都有严格的限制时间，具体为：

语言理解.....8 分钟      抽象推理.....25 分钟

概念类比.....22 分钟      空间推理.....25 分钟

数学能力.....22 分钟      机械推理.....22 分钟

再加上测验各部分之间衔接的时间，整个测验大约需要两个半小时。

测验有两种实施方式：纸笔作答和计算机施测。具体施测过程与前面介绍的测验类似。

## （四）测验的指导语及样题

### 1. 总指导语：

**本测验包括六个部分：语言能力测验、概念类比测验、数学能力测验、抽象推理测验、空间推理测验、机械推理测验。这些测验是进行职业选择的重要依据。**

**每项测验都有规定的时间。每项测验开始之前，有一段时间阅读该项测验的说明、例题，熟悉测验方式。每项测验由测验主持人发出指令后才可开始。每项测验若在规定的时间内提前完成，不可自行开始下一项测验，须等测验主持人的指令。若提前完成一项测验，可以检查该项测验，但不准检查、修改该项测验之前的其他测验答案。**

**答案必须填在答题纸上相应的地方。请用铅笔答题。不要在本手册上作任何草稿、记号。**

## 2. 样题

### （1）语言能力测验

**本测验考查语言表达能力。共有 10 个句子，每个句子被分为 1、2、3、4、5 五个部分，但不是按正常语序排列的。请找出理顺句子的最佳方案，并在答题纸相应的地方填出理顺后的各部分的相应序号的顺序。如：**

**经济的/人们/维持/希望/繁荣**

**1     2     3     4     5**

**句子理顺后应为：**

**人们/希望/维持/经济的/繁荣**

**故正确答案应为：2 4 3 1 5 应在答题纸相应的地方填“2 4 3 1 5”。**

**测验时间共 8 分钟，请尽量做到又快又对。如果没有把握，可以猜测。回答正确得分，答错不倒扣分。**

**答案请填写在专用的答题纸上，不要在本测验手册上作任何记号、草稿。**

### （2）概念类比测验

本测验考察语言推理能力。共有 50 个未完成的句子，每句有 5 个供选择填空的答案，请从中选出你认为最合适的一对词（第一个词填入句首，第二个词填入句末），使句子合理且完整。如：

…对于“水”就好比“吃”对于…

- ① 连接—动力
- ② 聊—散人
- ③ 喝—食物
- ④ 大碗—工业
- ⑤ 喝—散人

正确答案为：③。故应在答题卡相应的地方填“3”。

（3）数学能力测验

（4）抽象推理测验

本测验考察抽象推理能力。你将在每个题的图中看到两行图案或图形（见图 6-1）。第一行包括四个“问题图”，第二行包括五个“答案图”。四个问题图组成一个系列，从左向右具有一定的规律。请找出这个规律，并从答案图中选出一个图，它恰好符合问题图的规律，成为该系列的下一个图形（第五个图形）。例如，下一个例题的正确答案为“5”。

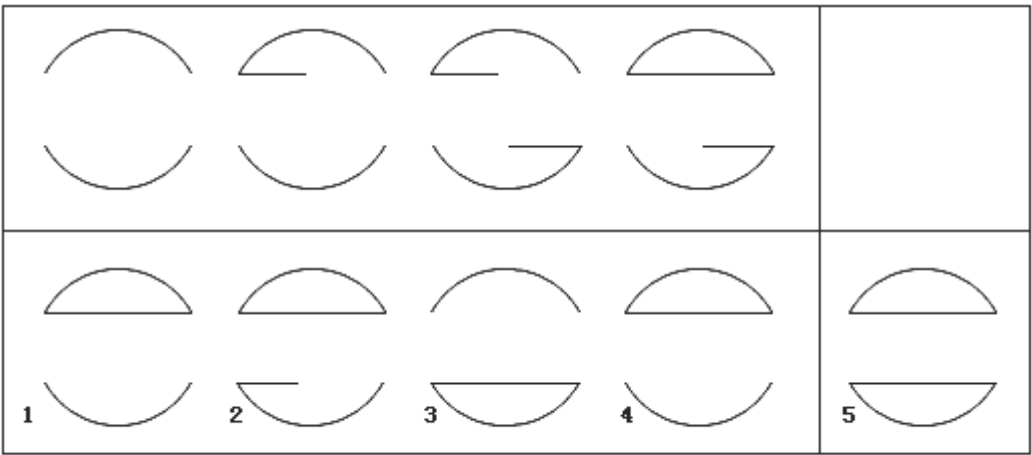


图6-1 抽象推理测验示例

（5）空间推理测验



本测验考察空间想象、推理能力。在每一个纸样子的右边，有四个几何图形。请确定哪一个几何构形可以由所给出的纸样子折叠而成（见图 6—2）。图示的纸样都是几何构形的表面。例如，下一个例题中答案为“3”。

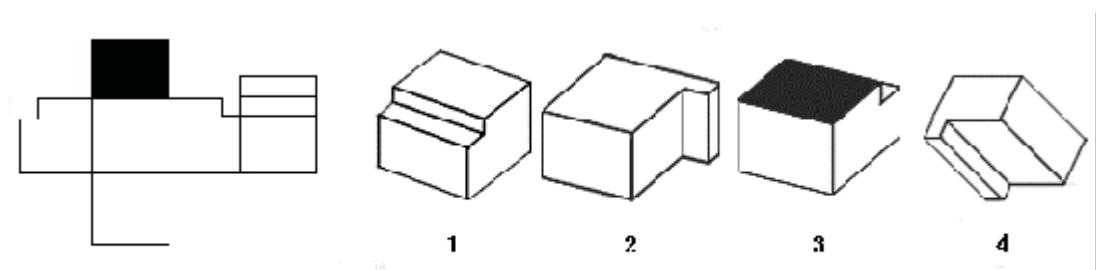


图6-2 空间推理测验

（6）机械推理测验

本测验考察机械推理能力，测验由关于机械原理一般常识方面的问题组成，共有 70 个题目。每个题目都有 1、2、3 三个选择；如果题中只有两种情况，那么当你认为 1、2 两种情况相同或均可或都不对时，就回答 3。注意：只能有一个最佳选择。例如，下面的例题中（见图 6—3），正确答案是“2”。

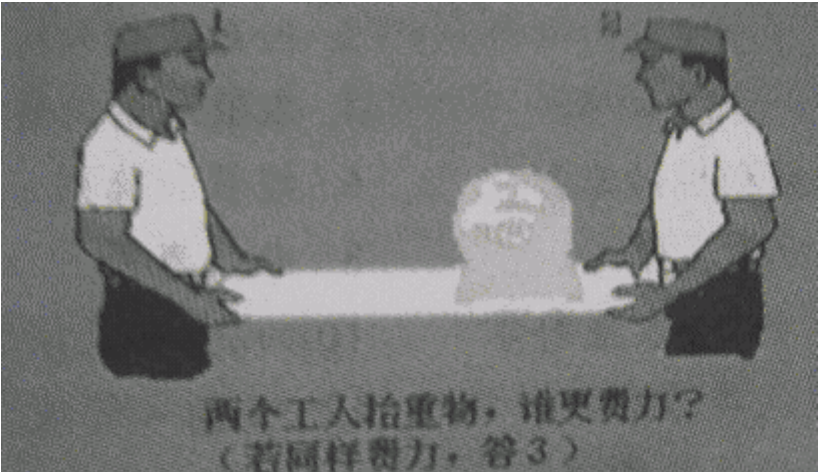


图6-3 机械推理测验

（五）报告样例

## **多项能力与职业意向咨询**

**姓名：\*\*\* 性别：女 年龄：18 学历：高中**

**各项能力评价：每个维度上的得分**

**各能力测验结果均为百分制，平均分为 55 分，40—70 分之间均属中等，71—90 分为优秀，91—100 分为极其优秀。**

**适宜职业排序结果：最适宜的为技艺和手工艺，最不适宜的为体育运动和娱乐。**

**职业类别：技艺和手工艺。**

### **（六）本测验需要特别注意的事项**

1. 多项能力与职业意向咨询可用于测查大、中学生和社会上一般人员的能力素质水平，不适合于管理人员的招聘和选拔。在测验使用上，这一点非常重要。
2. 多项能力与职业兴趣的结果可相互参照，因为能力加兴趣才是职业成功的重要保证。个体要根据自身特点进行协调，以使能力和兴趣有最大程度的重合。

## **第六章 能力测验**

### **第二节 数量分析能力测验**

#### **一、数量分析能力测验的目的与功能**

##### **（一）目的**

数量分析能力测验是对应试者对数量图表等信息的敏锐感和分析能力，以及数量分析能力进行测查评定。

##### **（二）功能**

本测验的测查不需要任何特别的专业知识背景，测查内容和范围包括：

1. 对数值、图表的敏感性；
2. 快速的综合分析能力；

3. 一定的快速数字估算能力。

## 二、数量分析能力测验的特点

（一）难度测验，有考试时间的限制；

（二）题型为单项选择题，每题有、且只有一个正确答案；

（三）测验由若干图表分析题构成，其中每组图表分析题内包含 1—3 个图表和与图表有关的 2—6 道单项选择题。要求应试者通过对图表显示内容的分析，对每道题做出正确选择。

（四）适用于有一定学历水平的人员。

## 三、数量分析能力测验的适用对象

本测验适于在中层管理人员选拔中作为能力考核的工具，适用于大学本科或以上学历或相当学历的应试人员。

## 四、数量分析能力测验的构成

测验包括 5 组图表分析题，每组图表分析题内包含 1—3 个图表和与图表有关的 2—6 道单项选择题，共计 25 道题。所有题目均为五择一的多重选择形式。要求应试者在 20 分钟内完成。鼓励应试者对不能完成的题目进行随机猜测，测验不进行倒扣分。

## 五、数量分析能力测验的施测过程

测验严格限定时间，测验时间为 20 分钟。纸笔或计算机作答均可。具体实施程序与前文介绍的测验类似。

## 六、测验的样题（含指导语）

指导语：

请您迅速阅读题目中的图表，找出关键信息，然后对所提问题作答。有些题目的备选项并不包含准确答案，请您通过估算找出最接近的答案。

样题（略）：类似于看数据表，然后做选择题。

## 七、报告样例

### **数量分析能力测验**

**姓名：\*\*\* 性别：女 年龄：25 学历：大本**

**测验的目的与功能：**

**具备一定的数量分析能力是高素质企业管理人员所必需的，也是现代企业对未来管理者的基本要求之一。管理人员数量分析测验调查内容和范围包括：**

- （一）对数值、图表的敏感性，**
- （二）快速的综合分析能力，**
- （三）一定的快速数字估算能力。**

**测验得分（百分等级）：92.8**

## 第六章 能力测验

### 第三节 管理人员逻辑推理测验

#### 一、管理人员逻辑推理测验的目的与功能

##### （一）目的

对中层管理人员来说，分析问题、解决问题的能力十分重要。本测验可以帮助企业选拔具有很强的语言分析能力，能迅速深入地加工信息，找到问题关键，并善于分析语言文字表达的信息，能基于事实而非主观臆断地做出判断的优秀管理人才。

##### （二）功能

测查应试者思维的准确、敏锐程度，逻辑推理的严密性和连贯性。

#### 二、测验的特点

（一）难度测验，有考试时间的限制。

（二）题型为单项选择题。

（三）测验由若干逻辑推理题构成，其中每道逻辑推理题内或者包含一段对前提假设的描述，要求应

试者根据假设推出合理的结论；或者提供一段对某一事件的结论，要求应试者在各选项中找出使结论成立的前提假设。

（四）为使应试者熟悉测验题型，正式测验前安排有练习时间。

（五）适用于有一定学历水平的人员。

### 三、适用对象

本测验适于在中层管理人员选拔中作为能力考核的工具，适用于大学本科或以上学历或相当学历的应试人员。

### 四、测验的构成

测验包括若干组逻辑推理题，每组逻辑推理题内包含一段描述性文字及 1—5 道单项选择题，每道选择题含有五个选项，所有题目均为五择一的多重选择形式，共 23 道选择题。要求应试者在 20 分钟内完成。鼓励应试者对不能完成的题目进行随机猜测，测验不进行倒扣分。

### 五、测验的施测过程

测验严格限定时间，测验时间为 20 分钟。纸笔或计算机作答，具体施测过程如下：

（一）依据预定的参试人数选择好适宜的测验地点，布置考场。考场环境应安静整洁，无干扰，采光照明良好。

（二）准备好测验所用的如下材料：测验题本、专用答题纸、铅笔、橡皮，保证每位应试者有以上完整的测验材料及用品。

（三）安排考生入场，并宣布测验注意事项。

（四）计时，注意监督应试者。

（五）测验时间到，回收题本和答题纸，测验结束。

### 六、测验的样题（含指导语）

指导语：

回答以下各题之前,请仔细阅读所给材料,它们包括几句话或一个段落,找出句中包含的逻辑关系,并选出答案。有的问题可能不止一个答案,请选出最佳答案。请不要假定与材料不符、不可能或多余的条件。

样题:

去年的通货膨胀率是1.2%,今年是4%。我们可以做出结论,通货膨胀处于上升趋势,明年将继续升高。

以下说法中,哪一句削弱了上述的结论?

(A) 人们从经济数据中,抽取了一个有代表性的样本,用以计算通货膨胀,而没有用所有获得的数据来计算。

(B) 去年石油价格下降,导致通货膨胀率暂时下降,因而低于目前稳定的4%。

(C) 部分工人的工资增加了,这直接导致通货膨胀率上升至4%或更高。工资上涨还会使通货膨胀率继续上涨。

(D) 去年1.2%的通货膨胀率是因为持续10年的低水平。

(E) 政府的干预不能显著影响通货膨胀率。

七、报告样例

### 逻辑推理能力测验

姓名:\*\*\* 性别:女 年龄:25 学历:大本

测验的目的与功能:

对管理人员来说,分析问题、解决问题的能力十分重要。逻辑推理测验可以帮助企业选拔具有很强的分析能力,能迅速深入地加工信息,找到问题关键,并善于分析语言文字表达的信息,能基于事实而非主观臆断地做出判断的优秀管理人才。逻辑推理测验主要测量应试者思维的准确、敏捷程度,逻辑推理的严密性和连贯性。

测验得分(百分等级):26.1

## 第四节 敏感性与沟通能力测验

### 一、敏感性与沟通能力测验的目的和功能

#### （一）目的

敏感性和沟通能力测验是为鉴别有潜力的营销人才，以及选拔和培训营销人员而编制的。

#### （二）功能

1. 本测验通过考察应试者对特定问题或现象的分析、处理的深度和把握问题实质的敏锐程度，以及在面对面营销情境下的应对方略，了解应试者在沟通中把握人际信息的敏感性及其对事物的洞察、预见和分析能力。

2. 同时可以了解其沟通策略模式，预测其在推销中说服、影响、感染他人并且达成协商成果的能力。

3. 基于对应试者在开放式沟通倾向、操纵式沟通倾向以及营销意识、营销常识四方面的量化分析，可全面评估其沟通技巧和有效程度，为其有针对性地修正和提高人际沟通技巧提供依据。

### 二、敏感性与沟通能力测验的特点

本测验分两部分：敏感性测验与沟通能力测验。

（一）敏感性测验测查应试者对人事物的洞察、分析和预见能力，特别是在认识和把握问题的实质并进行分析处理时敏锐地捕捉人际信息、利用人际关系有效地解决问题的能力；

（二）沟通能力测验侧重于对应试者在营销情境中运用人际沟通技巧和策略方面的考察。

### 三、敏感性与沟通能力测验适用对象

营销人员、面临职业选择的毕业生和其他人员、企业应聘人员、企业在职人员。

### 四、敏感性与沟通能力测验的构成

本测验的题目多数采用设定特定情境下的具体问题的形式，提供的备选答案是针对上述问题的不同见解或处理办法，被试需按题中指定的身份进行角色扮演，选择出自己认为最恰当的选项。

测验由 90 道题组成，分为敏感性测验和沟通能力测验两部分，每一部分各有 45 道题，每个问题有四个答案供选择。第二部分分成两个分测验：沟通倾向与营销常识。测验有以下维度：

（一）敏感性

（二）沟通行为倾向：

1. 开放式沟通倾向
2. 操纵式沟通倾向
3. 非沟通倾向（营销意识）

（三）营销常识

## 五、测验的施测过程

测验分两部分，第一部分敏感性测验，第二部分沟通能力测验，每一部分限时 30 分钟完成。

两部分测验可合成一套完成，先完成敏感性部分（30 分钟），再做沟通能力测验（30 分钟），中间可安排休息时间（5 分钟），测验时间约为 65 分钟左右。两个部分的测验也可单独作为独立的测验使用，测验时间限定为 30 分钟。

纸笔或计算机作答，可集体施测。具体过程同前。

## 六、测验样题

（一）敏感性测验部分

样题：

**你是某儿童团体的指导员，有一个儿童，不服从你的领导和忠告，因此破坏了这个团体的良好风气。在这种情况下，你应该怎么办？**

- ①把这个儿童转到其他社团。**
- ②与社团中其他儿童商讨这一问题。**
- ③和这个儿童商讨这个问题。**
- ④把这件事告诉他的父母。**

**在以上提供的四个答案中，如果你认为③是最适当的答案，就在专用的答题纸上涂黑相应题号后的③。**

**每个题目只选一个答案。**



## （二）沟通能力测验部分

样题：

**你是一位服装销售员，你已经花「半个小时」拿「好几套衣服」给一个顾客看，然后他突然说“我想看看别的商店再作决定”，在这种情况下，你应该怎么办？**

**①使他知道你对买卖做不成感到失望。**

**②问他现在是否想买，假如真的想买，再多拿几套给他看。**

**③告诉他不必到别的店去看，因为你们店里的货品都是最好的。**

**④告诉他你非常高兴他多看几家商店，详细考虑后再来购买。**

**在以上四个答案中，如果你认为④是最适当的答案，就在专用的答题纸上涂黑相应题号后的④。**

**每题中只选一个你认为最适当的答案。**

## 七、维度定义

（一）敏感性（S）。指对人际事物的洞察、预见和分析能力。敏感性高的人善于从纷繁复杂的事物中看到内在的逻辑联系，敏感地把握问题所在的关键的人际信息，合理地通过角色扮演有效地解决问题，特别是人际关系问题。高敏感性是具备良好人际沟通能力的前提。

（二）沟通行为倾向。沟通倾向的适当与否代表一种能力，指从事营销活动的人员在与营销对象进行沟通中采取有效方式和策略的能力。这种能力多表现为营销情境中人际沟通的策略性、技巧性、不同个体因多种因素影响形成不同的行为模式和策略倾向，我们称为沟通行为倾向，区分为开放式沟通倾向、操纵式沟通倾向和非沟通倾向，定义如下：

1. 开放式沟通倾向（K）。在沟通中非常讲究策略，善于接纳别人的观点，能很好地理解别人的立场，能有技巧地向他人提出建设性意见，温和而有说服力地说服他人；不生硬或过分地表露自己的想法，能在个人和他人的价值观之间进行很好的平衡，善于在不同意见之间周旋，往往能获得协调的、双方欣然接受而满意的结果。

2. 操纵式沟通倾向（C）。在沟通中策略性不强，以自我为中心，往往表现出对别人立场、观点的不

理解，而只从自己的立场出发刻意推销自己，可能十分健谈，但比较生硬，不善营造良好的沟通气氛，不能为自己在不同意见间的周旋留下足够空间；由于明显地表露自己的推销意图，使他人有被操纵感，往往不能达到有效的沟通目的。

3. 非沟通倾向/营销意识（F）。在沟通中缺乏营造良好沟通气氛的意识，对各种有关营销的机遇敏感性不高，不会主动积极地寻找和把握交流沟通的机会，对问题的处理缺乏人际沟通取向的考虑，而往往采取非沟通途径的做法。

开放式沟通倾向是有效沟通模式，是正性的，在测验中该分数作为正面评价的指标；操纵式沟通倾向和非沟通倾向则对达到沟通目的有负性影响，这两项分数作为参考分从负面给予描述。

（三）营销常识（I）。营销常识是关于营销实务各类活动的一般原则、方法及某些普遍观念。营销常识是指营销人员应付各种营销情境、实现高绩效的必要知识储备，与个体的教育背景和业务经验有关。

## 八、报告样例

### 敏感性与沟通能力测验

#### **敏感性：**

对人事物有很好的洞察、预见和分析能力，能从纷繁复杂的事物中看到内在的逻辑联系，能够敏感地看到人际关系的微妙信息，迅速把握并找到行之有效的解决问题特别是处理人际关系的方案，方式、方法恰当。

#### **沟通技巧：**

具有突出的开放式沟通风格，在沟通中非常讲究策略，善于接纳别人的观点，能很好地理解别人的立场，能有技巧地向他人提出建设性意见，温和而有说服力地说服他人，不生硬或过分地发表自己的想法，能在个人和他人的价值观之间进行很好的平衡，善于在不同意见之间周旋，并总能获得协调的、双方欣然接受而满意的结果。

#### **营销意识：**

能够敏感地把握各种有关营销的机遇，并能迅速把握时机，做出适当反应，具备营销人员的素质，在业务中常能意识到维护作为营销人员的职业形象感。

#### **营销常识：**

具有完善的知识结构，具备系统而全面的推销业务常识，在复杂的业务情境中有准确地把握信息、迅速地积累经验的潜能，在业务水平上有较大的发展可能。

### 敏感性和沟通能力测验结果

姓名：\*\*\* 性别：女 年龄：25

1	2	3	4	5
敏感性	开放式沟通倾向	封闭式沟通倾向	营销意识	营销常识
92	97	-7	-4	92

#### 九、本测验特别需要注意的事项

（一）敏感性和沟通能力两部分测验可作为一整套测验使用，也可以作为独立测验分别使用。但由于人际敏感性是具备高超人际沟通能力的前提，能得到两个分数可对应试者的发展潜能和绩效做综合预测，建议两部分测验都要做。

（二）人际敏感性是较有普遍性的心理品质之一，除营销人员之外，很多行业也需要强调从业者的个人适应力、人际敏感度，这时应考虑使用敏感性测验。

（三）本测验的测评结果可为营销人员下一步的培训提供建议，用于有的放矢地设计培训课程，获得良好的培训效果。

（四）营销意识的强弱由维度非沟通倾向（F）的分值大小而定，同时参考开放式与操纵式绝对分值的差额。一般上述差额越大，F 值越小，沟通有效性越大，营销意识越强。



职业经理 MBA 整套实战教程

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第七章 组织人事调查

### 第一节 个体行为评估

#### 一、工作感觉评定

##### （一）工作感觉评定的功能和特点

##### 1. 目的

调查员工对现任工作的满意度，为组织了解员工的工作态度，预测其工作效率和稳定性提供信息。

##### 2. 功能

工作满足感是“员工对于他们工作的一般性意见”。工作满足感可影响员工的工作绩效，比如影响生产力，旷工、离职行为的发生。通过调查员工的工作感觉，可以了解员工对工作的满意状态，从一个侧面反映了组织内部的管理和分工的合理和有效程度，为管理效能诊断和组织发展、变革提供重要信息。

##### 3. 特点

员工对于环境的知觉，才是其行为的真正直接基础。因此，个体知觉是决定工作满意度的极其重要的

因素。因此，直接询问员工感受的问卷调查法是最为有效的测量工具。

## （二）工作感觉评定的应用

### 1. 适用对象

本调查广泛适用于各行业、各层次人员，特别适于对需要进行管理诊断、团队建设，以及处于组织变革、发展阶段的特定单位和部门的全体人员进行集体测查。

### 2. 调查问卷的构成

问卷由 18 个题目构成（更复杂一些的问卷可能由 40—50 个问题构成），每个题目都是关于对现任工作看法的描述，包括行为上、认知上和感情上三方面。

### 3. 调查的施测过程

这个调查需要大约 5 分钟。具体的施测过程与前面所讲的测量工具类似。

### 4. 调查的样题（含指导语）

指导语：

**“每个人都对自己的工作有一定的感受或评价。请阅读下面每一个问题，并在题目右边的选择中圈出您的感受或意见。请尽快回答，不要迟疑。”**

**1-完全同意，2-同意，3-没意见，4-不同意，5-极不同意。”**

样题：

**（1）我的工作经常会有足够的趣味，不会令人厌烦。 1 2 3 4 5**

### 5. 调查结果的使用建议

通过工作感觉调查可了解员工整体的和个体的工作满意水平，如果希望进一步影响和改善员工工作满意度，可依据调查结果，结合本单位或部门的实际情况，从以下决定工作满意度的因素中分析筛选出关键的要点。决定工作满意度的因素分布在以下各个层面上：

#### （1）个体层面：能力、性格、学习、价值观

(2) 工作相关层面：挑战性的工作、公平的待遇、支持性的环境

(3) 团体层面：团队内部的沟通状况

(4) 组织层面：组织结构、人力资源实务、绩效考核与激励、组织文化、工作压力

## 第七章 组织人事调查

### 第一节 个体行为评估

#### 二、价值取向评估

##### (一) 价值取向评估的功能和特点

##### 1. 目的

调查员工在理论、经济-政治、唯美和社会四种价值取向的相对强度。可由此鉴别职业志向，评定工作动机。

##### 2. 功能

通过价值观评定，一方面企业管理者可了解自己的价值取向，并把自己的价值取向正确而合理地用到企业的日常管理活动中去；另一方面，可揭示员工价值取向构成模式，为培养员工多元化兴趣、建设和发展企业文化提供依据。

##### 3. 特点

价值观是一种基本信念，它带有判断的色彩。基于价值观影响个体态度和行为的以上逻辑思路，该调查采用问卷形式，让受试者评定对代表不同价值追求的多种活动的好恶情况，依据应试者的好恶程度来确定价值观的强度排序，从而可揭示个人的价值观体系。

##### 4. 维度定义

不同价值取向的人有不同的追求：

(1) 理论取向。重视事情的真理，一般不拘于人情。追求逻辑和理性，强调经由批判而理性的方法发掘真理。

(2) 经济—政治取向。注重功利，追求实用性，讲究经济效益，强调权力的取得与影响力，认为人生的物质享受比精神享受更重要。

(3) 唯美取向。追求艺术美感，强调形式与和谐性。追求丰富多彩的生活，追求美的世界。

(4) 社会取向。重视人际关系，讲究博爱、平等、公平和友谊。

## (二) 价值取向评估的应用

### 1. 适用对象

适用于各个行业、各个层次的人员。特别是处于变动阶段的组织。

### 2. 调查问卷的构成

每个维度 20 个条目，共有 80 个条目。

### 3. 调查的施测过程

该调查大约需要 20 分钟。

### 4. 调查的样题（含指导语）

指导语：

**“人的兴趣爱好是有所不同的。请阅读下面的每一个陈述或情形，并判断是否符合你自己的情况。把圈的答案圈在后面。请尽快回答，不要拖延。”**

样题

**(1) 喜欢结识新朋友。 是\_\_\_\_\_否\_\_\_\_\_**

### 5. 调查结果的使用建议

调查全体员工的价值观测量结果也很有价值，它可揭示团队或组织中员工的价值追求层次，为有针对性地丰富和建设企业文化，提高员工对组织的认同和忠诚度，实现更好的团队绩效提供依据。

## 第二节 领导行为评估

### 一、沟通方式评定

#### （一）沟通方式评定的功能和特点

##### 1. 目的

调查管理者是否了解正确的上下沟通的知识、概念和技能。

##### 2. 功能

沟通方式评定可反映管理层对正确的人际沟通的基本概念和技能的掌握程度，考察是否能够正确处理组织中的人际关系。

##### 3. 特点

本调查选取代表涵盖面最大的向上、向下和横向沟通三个层面的一些具体情景，这些情景或者比较尴尬，难于应付，或者存在着潜在的利益冲突，都有一定的难度，以此考察被调查者是否有正确处理的意见，从而反映其掌握沟通技能、概念的程度和综合应用有效沟通原则的水平。

#### （二）沟通方式评定的应用

##### 1. 适用对象

适用于管理者，特别是中层管理者。

##### 2. 调查的构成

该调查由 12 个题目构成，每个题目有三个选项。题目是一些情境性的问题。

##### 3. 调查的施测过程

该调查大约需要 15 分钟。

##### 4. 调查的样题（含指导语）

指导语：



**“每个人都有独特的与人沟通、交流的方式。阅读下面的情境问题，选择出您认为最合适的处理方式，圈出相应的字母。尽快回答，不要犹豫。”**

样题：

**你的上司的上司邀你共进晚餐，回到办公室后，你发现你的上司很好奇，此时你会：**

- A. 告诉他详细内容。**
- B. 不透露蛛丝马迹。**
- C. 趣味描述，淡化内容的重要性。**

#### 5. 调查结果的使用建议

通过测查中层管理人员的沟通方式，可了解该层次人员的沟通技能水平，在一定程度上可预测团队绩效，同时为监控和提高中层管理绩效提供线索，为设计合理的培训和发展计划提供建议。沟通在团体或组织中有四项主要功能，参照比较特定组织在实现这四项功能上的表现，可作为进一步考察组织管理者的沟通效能水平的一个思路。这四项功能是：

（1）资讯沟通——当个人或团体需要作决策时，或者组织中各团体或团体中各成员间通过运作来完成任务时，通过沟通才能得到必要的信息，以分析、选择和执行各种方案。

（2）情感表达——对各层次员工而言，他们的工作团体往往是他们跟别人建立社交互动关系的主要来源。

（3）激励——沟通也是激励员工的手段。

（4）控制——以沟通来控制成员的行为。

## 第七章 组织人事调查

### 第二节 领导行为评估

#### 二、冲突应付方式评定

##### （一）冲突应付方式评定的功能和特点

### 1. 目的

测查应试者在各种冲突情境中解决问题的方法，从而反映出其应付冲突的方式和风格。

### 2. 功能

冲突处理方式可区分为两个向度：合作性和肯定性，经过不同的组合，构成五种类型：竞争、统合、退避、顺应、妥协（参看图 7—1）。每个人都有自己潜在惯用的冲突处理方法。通过该调查，应试者可以确定自己的冲突处理风格，从而有效地指导和修正自己的行为。

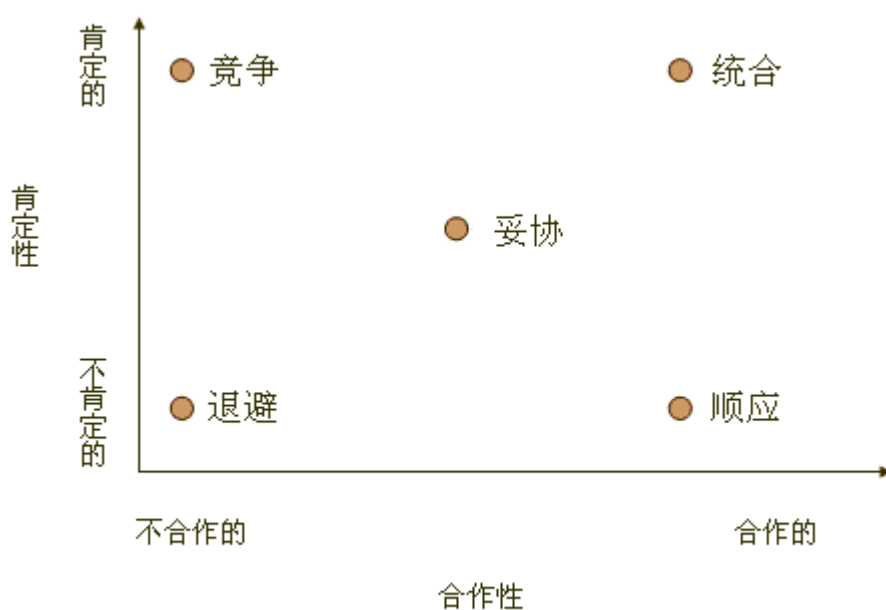


图7-1 冲突处理方式

### 3. 特点

调查采用自评式问卷形式，选取代表三大类冲突处理风格，即非抗争型（退避与顺应）、解决问题型（统合与妥协）和抗争型（竞争）的典型行为方式和态度建构而成，有很大的概括性。

### 4. 维度定义

（1）合作性：指某方试图满足对方需求的程度。

（2）肯定性：指某方试图满足自己需求的程度。

（3）竞争（肯定的与不合作的）：当一个人只顾追求自己的目标和获取利益，而不顾虑冲突对方的

影响时，此行为为竞争或支配。

（4）统合（肯定的与合作的）：当冲突的双方都希望满足对方的需求时，便会合作而寻求两者皆有利的结果。

（5）退避（不肯定的与不合作的）：指一个人承认冲突的存在，却采取退缩或压抑的方式。

（6）顺应（不肯定的与合作的）：当一个人希望满足对方时，可能会将对方的利益摆在自己的利益之上。为了维持彼此的关系，某一方愿意自我牺牲。

（7）妥协（中等肯定与中等合作）：当冲突双方必须放弃某些东西时，则会因为分享利益而导致妥协的结果。

## （二）冲突应付方式评定的应用

### 1. 适用对象

适用于管理者，特别是中层管理者。

### 2. 调查的构成

调查由 30 个题目构成。每个题目的答案由 1——7 的等级评定构成。

### 3. 调查的施测过程

调查大约需 5—8 分钟。

### 4. 调查的样题（含指导语）

指导语：

**“每个人都有自己应付冲突的方式、风格。阅读下面的题目，用 1-7 分的等级来评定它对你的符合程度。”**

**“1”=总是如此，“7”=从不如此，其他分数表示中间的程度，分数越小越符合你，分数越大越不符合你。**

样题：

**(1) 我和上司出现分歧时，我会以折衷的方式解决。 1 2 3 4 5 6 7**

5. 调查结果的使用建议

作为管理者应了解各种冲突处理方式的特点，适合于处理何种情景的冲突，如缺乏这方面的知识，应注意学习和补充。五种冲突解决方式的最适合使用情景如表 7—1 所述。

**表 7—1 五种冲突处理取向的适用时机**

<b>冲突处理取向</b>	<b>适用情景</b>
<b>竞争</b>	特别需要快速，有决断的行动时。 在重要的争议上，需要介入不寻常的行动时。 与组织的利益休戚相关，而且确定自己正确无误时。 反对因采取非竞争性行为而获益的人时。
<b>整合</b>	当双方的考虑非常重要又无法妥协，需要整合性的解决办法时。 当目的是要学习时。 想要合并不同的观点时。 想要合并双方的考虑达成共识，以取得承诺时。 想要让关系免受干扰，以利于解决问题时。
<b>回避</b>	问题很琐碎，又有其它问题迫近时。 当觉察到需要无法满足时，当前关系破裂后的害处超过问题解决后的利益时。 想要冷静下来重新考虑时。 当新获得的信息取代了刚才的决定时。 当另有他人能够更有效的解决冲突时。 当争议回避，或有引发其他争议的征兆时。
<b>顺从</b>	当发现自己犯错的时候——让自己有较好的立场，要学习或体现自己是讲理的时候。 当争论的主题对他人更重要时——满足他人，保持合作。 当为了寻求建立社会性信用时。 当技不如人或有所损失时，使损失降至最低。 当事情与确定是特别重要之时。 让部属从错误中学习成长之时。
<b>妥协</b>	当目标很重要，但不值得更努力争取时。 当实力相当的对手承诺各自独占目标时。 在时间压力下达到取直之计时。 在整合与竞争都失败时。

## 第二节 领导行为评估

### 三、工作习惯评定

#### （一）工作习惯评定的功能和特点

##### 1. 目的

调查管理者的官僚倾向，即是否能接受或偏好严格的制度和规范化模式。一般来说，偏好过强或过低均不适宜。

##### 2. 功能

工作习惯评定可评估个人对工作稳定性的需要，了解对制度化工作模式的态度和对企业制度的接受或习惯程度。

##### 3. 特点

本评定以科层结构的概念为基础，围绕科层结构的以下特征来构建，概括了大多数机械式组织结构模式的特征，也反映了高度复杂化和正式化的典型特征。

#### （二）工作习惯评定的应用

##### 1. 适用对象

适合于管理者，也适合于全体员工。

##### 2. 调查的构成

本调查由 20 个题目构成。

##### 3. 调查的施测过程

本调查大约需要 20 分钟。施测的步骤与之前所讲测量工具类似。

##### 4. 调查的样题（含指导语）

指导语：

**“每个人都对环境和工作方式有自己的态度，形成一定习惯。请判断下列情况是否与您相符合，并在题目后面相应的横线上画圈‘0’，请尽快回答，不要拖延。”**

样题：

	基本	基本
	肯定	否定
<b>(1) 我重视工作的稳定性。</b>	_____	_____
<b>(2) 我喜欢有制度的公司。</b>	_____	_____

## 5. 调查结果的使用建议

通过本调查的结果可了解员工对组织结构复杂化和正式化程度的接受、适应程度，一方面可反映员工的潜在创造力和企业活力，另一方面也可作为企业了解其组织结构状态与组织发展现状、组织目标吻合程度的一个手段。

组织结构的形成和发展有两种模式，一是权变模式，指组织结构随情景因素、企业经营策略、组织规模、技术例行性程度、外界环境的不确定性的变化而变化；另一种是非权变模式，指组织的掌权者为了谋私利，会选择一种能维持或提高他对组织控制的结构设计。这种组织的内部权力结构很少会起变化，组织结构相当稳定。对于特定企业，如果能确定其组织结构产生于何种模式，对分析和预测员工绩效和组织发展有很大帮助。

## 第七章 组织人事调查

### 第二节 领导行为评估

#### 四、变革意识评估

##### （一）变革意识评估的功能和特点

##### 1. 目的

测查管理者对事物变化的敏感性、接纳性，以及对待事物的灵活性和创新意识。

## 2. 功能

考察变通性，尤其是打破现状而求变化的变革意识。变革意识是组织能否长期生存的根基，需要做经常、持久的工作来加以强化。

## 3. 特点

通过考察被调查者对事物看法的广度和所持态度的固执程度来了解其变通性。

### （二）变革意识评估的应用

1. 适用对象：适合于企业员工。主要针对管理人员，特别是中层以上的管理人员。

2. 调查的构成：本调查包含 20 个题目，答案是包含 1—6 的程度判断。

3. 调查的施测过程：本次调查大约需要 10 分钟，其他步骤与之前所讲测量工具类似。

4. 调查的样题（含指导语）

指导语：

**“每个人对自己的工作和环境有不同希望，请阅读下列陈述，用 1—6 的等级评定每个陈述对你的适宜程度，在后面圈出你的选择。”**

样题：

**（1）大多数人并不知道什么才是对他有益的。 1 2 3 4 5 6**

## 5. 调查结果的使用建议

培养适宜开发变革意识和创造性的组织氛围，应在以下五个方面创立充分条件：

- （1）保持良好的开放式沟通。
- （2）创造良好的团结合作气氛。
- （3）从事开发创造活动的人员与操作人员适度分离。
- （4）容许创造性研究人员有一定的弹性时间。

(5) 切忌过早下结论和批评。

## 第七章 组织人事调查

### 第三节 团队行为评估

#### 一、团队健康度测试

##### (一) 团队健康度测试的功能和特点

##### 1. 目的

帮助人们了解自己团体的现状。

##### 2. 功能

一个团队默契而优秀的配合，是它创造卓越成效的关键。本测试从共同领导、团队工作技能、团队气氛、团队凝聚力、成员贡献水平等五个方面来描述团队现状，从而指导人们更有效地进行团队建设和团队学习。

##### 3. 特点

调查以配合默契的优秀团队在上述五个方面的表现为标准，受测团队的状态越接近这一标准，可以说越健康，预测团队绩效会越好。

##### 4. 维度定义

(1) 成员共同领导的程度：指一个团体的每一个成员都可以并有义务分享一份领导责任，一个团队是大家共同来领导的。如果一个团队是独裁的，那它的健康程度就低。

(2) 团队工作技能：指成员在一起工作相处的技巧。

(3) 团队氛围：指团队成员共处的情绪和谐度和信任感。

(4) 团队凝聚力：指团队成员对目标的一致性。

(5) 团队成员的贡献水平：指团队成员为实践自己的责任所付出的努力和成就程度。

##### (二) 团队健康度测试的应用



### 1. 适用对象

接受诊断的团队内部的成员。

### 2. 调查的构成

本次调查包括 25 个题目，共分为 5 个维度，每个维度 5 个题目。答案为 1—4 分评定量表。

### 3. 调查的施测过程

与之前所讲测量工具类似。

### 4. 调查的样题（含指导语）

指导语：

**请用 1—4 分评定下列陈述是否符合你所在的团队。请尽快回答，不共寒温。**

**1-不适合，2-偶尔适合，3-基本适合，4-完全适合。**

样题：

**（1）每个人都有同等发言权并得到同等重视。** \_\_\_\_\_

### 5. 调查结果的使用建议

每一个维度的最高分为 20 分，得分越高越好。如果一个团队在这五方面都很出色，那它就会是一个优秀的团队，也必定是一个高绩效的团队。比较所在团队不同方面的得分，就可以粗略地了解自己团队的长短之处。

如果让所在团队的每一个成员都完成这一评定，就可得到：

（1）团队成员对团队的总体（平均化）的评价；

（2）可比较总体评价和每一个团队成员的评价，了解每一个人与其他人的看法的差距。

## 第七章 组织人事调查

### 第三节 团队行为评估

## 二、团体绩效评定

### （一）团体绩效评定的功能和特点

#### 1. 目的

评估团体在完成任务的团队互动过程中协调、配合的气氛，团体内部的凝聚力、团体工作技巧以及成员与领导互动模式的有效性。

#### 2. 功能

通过揭示团队成员互动运作机制的有效程度，评估团队绩效，据此结果可有的放矢地进行团队建设，有效提高成员士气，形成和保持良好的团队绩效水平。

#### 3. 特点

调查以普遍的高绩效团体在互动运作和配合中出现的典型行为表现、活动状态和气氛特征为标准，通过自评问卷的形式了解具体某个团队的运作状况与标准的接近程度，以此来评估该团体的绩效水平。

### （二）团体绩效评定的应用

#### 1. 适用对象

受诊断的团队成员。

#### 2. 调查的构成

本调查包括 20 个题目。均为基本肯定与基本否定的判断题目。

#### 3. 调查的施测过程

本次调查大约需要 10 分钟。

#### 4. 调查的样题（含指导语）

指导语

**“请从以下各方面评定您所在的工作团体。给出您的评定结果，在相应的横线上画圈‘O’，请尽快回答，不要遗漏。”**

样题：

基本 基本  
肯定 否定

(1) 气氛显得轻松愉快。

—— —

## 5. 调查结果的使用建议

本调查是团体绩效的检查工具，通过调查可获得有关团体绩效水平的信息。如希望进一步讨论影响团体绩效的因素，可参考以下思路：

(1) 团体所面临的外部条件会影响团体的绩效，如组织策略、职权结构、甄选程序、薪酬制度等，这些因素决定了外在条件对该团体的支持程度。

(2) 团体潜在的绩效水准绝大部分决定于成员的素质。

(3) 高度不确定性任务需要处理更多的资讯，更依赖团队的配合，如果团体沟通不良、领导不佳、内部冲突迭起，会影响团队绩效。

(4) 团体凝聚力在不同团体规范的作用下对团体绩效的影响不同。



职业经理 MBA 整套实战教程

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第八章 基于情境的测验

### 第一节 公文筐测验

#### 一、什么是公文筐测验

##### (一) 公文筐测验的概念

1. 公文筐测验 (in-tray test)，又叫文件筐测验，是让被试在所安排的假定的情境中扮演某种管理者

的角色，对事先设计的一系列文件进行处理，进而针对被试处理公文的方式、方法、结果等进行评价。

2. 公文筐测验是一种情境模拟测验，是对实际工作中管理人员掌握和分析资料、处理各种信息，以及做出决策的工作活动的一种抽象和集中。测验在假定情境下实施。该情境模拟一个公司所发生的实际业务、管理环境，提供给受测人员的信息包括涉及财务、人事备忘录、市场信息、政府的法令公文、客户关系等十几份甚至更多的材料。这些材料通常是放在公文筐中的，公文筐测验因此而得名。

## （二）公文筐测验的优点

1. 具有灵活性，可以因不同的工作特性和所要评估的能力而设计题目。
2. 作为一种情境模拟测验，它可以对个体的行为做直接的观察。
3. 由于把人置于模拟的工作情境中去完成一系列工作，为每一个被试都提供了条件和机会相等的情境。
4. 它能预测一种潜能。这种潜能可使人在管理上获得成功。
5. 由于公文筐测验能从多个维度上评定一个人的管理能力，它不仅能挑选出有潜力的管理人才，还能训练他们的管理与合作能力，使选拔过程成为培训过程的开始。
6. 在实践中，公文筐测验除用作评价、选拔管理人员外，还可用于培训，提高管理人员的管理技巧、解决人际冲突和组织内各部门间的摩擦的技巧，以及为人力资源计划和组织设计提供信息。

## （三）公文筐测验的取材

1. 在测验材料的设计上，主要围绕管理者的能力取材。管理者（这里特指组织领导者）的管理能力主要来自三个方面：自身素质基础、社会实践体验、所掌握的有关知识。管理能力的水平和发展取决于以上三个方面的交互作用和整合的结果，故管理能力是复合性能力。如果以偏重知识性的、或经验性的、或智力性的具体能力为主要测评内容，则难以保证较好的评价效果。
2. 管理者的有关知识，特别是有关的管理技术知识和业务性知识，虽然对现实管理能力有较大影响，但不作为公文筐测验的主要测评内容。其主要理由是：

（1）管理者的知识水平可以通过其他简便有效的办法评价；

(2) 知识水平在一定程度上易于通过培训、锻炼等形式提高;

(3) 知识欠缺的弊端一般可以通过其有效的管理活动弥补。

管理活动与人的认知思维活动一样复杂。因此,在目前科学发展水平的条件下,有现实意义的是对管理者应具备的基本能力水平进行结构化的测评,它可以为管理人才的评价提供比较有效的客观依据,并加深我们对管理能力本身的认识。

## 第八章 基于情境的测验

### 第一节 公文筐测验

#### 二、公文筐测验的应用

##### (一) 公文筐测验的功能和特点

##### 1. 功能

针对高层管理者的胜任要求,考察计划、授权、预测、决策、沟通等方面的管理能力,特别是考察综合各类业务信息、审时度势、全面把握、处变不惊、运筹自如的素质。

##### 2. 特点

(1) 情景性强。完全模拟现实中真实发生的经营、管理情境,对实际操作有高度似真性,因而预测效度高。

(2) 综合性强。测验材料涉及到经营、市场、人事、客户及公共关系、政策法规、财务等企业组织的各方面事务,考察计划、授权、预测、决策、沟通等多方面的能力,从而能够对高层管理者进行全面评价。

##### 3. 维度定义

公文筐测验所要测评的能力定位于管理者从事管理活动时正确处理普遍性的管理问题,有效地履行主要管理职能所具备的能力。一般来说,要考察以下五个维度:

(1) 工作条理性。

设计一定的任务情境和角色情境，要求被试判断所给材料的优先级。得分高的被试能有条不紊地处理各种公文和信息材料，能根据信息的性质和轻重缓急对信息进行准确地分类，能注意到不同信息间的关系，有效地利用人、财、物、信息资源，并有计划地安排工作。

## （2）计划能力。

得分高的被试能非常有效地提出处理工作的切实可行的方案，主要表现在能系统地事先安排和分配工作，识别问题及注意不同信息间的关系，根据信息的不同性质和紧迫性对工作的细节、策略、方法做出合理的规划。

评价计划时，在某种程度上要关注被试对其行为未来后果的考虑。例如，考察他们解决问题时是否考虑时间、成本、顾客关系或资源。计划也包括为避免预期的问题所采用的步骤，以及出现这些问题时，他们对问题的操作步骤与方法。

## （3）预测能力。

得分高的被试能全面系统地考虑环境中各种不同相关因素，对各种因素做出恰当的分析，并做出合乎逻辑的预测，同时对预测能提出行之有效的实施方案。

该维度包括考察三部分内容：预测的质量、所依据的因素、可行性分析。

评价预测时，要考察被试为了作出预测而利用公文筐内材料的程度，即是否综合各种因素做出分析。

## （4）决策能力。

该维度得分高的被试对复杂的问题能进行审慎的剖析，能灵活地搜索各种解决问题的途径，并作出合理的评估，对各种方案的结果有着清醒的判断，从而提出高质量的决策意见。

该维度包括考察三部分内容：决策的质量、实施的方案、影响因素。

## （5）沟通能力。

要求被试设计公文，撰写文件或报告，用书面形式有效地表达自己的思想和意见。根据评估内容，考察被试的思路清晰度、意见连贯性、措辞恰当性及文体相应性。

得分高的文章要求语言非常流畅，文体风格与情境相适应，能根据不同信息的重要性来分别处理，结

构性很强，考虑问题很全面，能提出有针对性的论点，表现出熟悉业务的各个领域。

## （二）公文筐测验的应用

1. 适用对象：中、高级管理人员。
2. 测验的构成：一是测验材料，二是答题册。
3. 测验的施测过程：基本过程同前。

注意：（1）严格计时，分测验也有时间限制。

（2）不但看结果，还要看处理方式。

（3）参试人数尽量不要太多。

4. 测试的样题（含指导语）

总指导语：

**这是一个“公文筐”测验，它模拟实际的管理情境，请件处理商业信息、文件和管理人  
员常用的信息。**

**这个模拟的具体假设情境是：**

**件是瑞克有限公司的市场营销部经理。你叫“王意峰”。**

**今天的日期是XXXX年2月8日、星期二。**

**现在的时间是：上午7点40分。**

**件刚刚来到办公室，正独自坐在办公桌前。今天早些时候，公司国际业务部总裁打电话  
通知件，公司的总经理已经辞职离开了公司。**

**这里为件准备了件今天需要处理的全部材料，放在专用的题料文件夹里。**

测验材料样例：

### 关于增加人事干部编制名额的请求

总经理：

经董事会批准，今后总公司、分公司两级的干部培训工作由人事部门全面负责。但是，在公司最初确定人事部门人员编制时没有培训工作这项任务。为了做好这项工作，需要给人事部门增加必要的编制名额，建议给人事处增加 3 人，每个分公司增加 1—2 人。

关于人事处增加的 3 个编制名额，请总经理审批；关于给分公司增加的编制名额，请批准各分公司从现有名额中调剂解决。

以上请示当否，请批示。

### 5. 报告样例

#### 公文筐测验

考号：101 姓名：\*\*\* 性别：男 年龄：22

1	2	3	4	5
工作条理性	计划能力	预测能力	决策能力	沟通能力
64.3	86.4	48.4	96.6	78.5

**工作条理性**——工作非常有条理，能分清事务的主次，并能据此安排自己的工作步骤，使工作按既定的次序进行，表现出一定的计划性。

**计划能力**——对工作的处理得当，分析能力较强，能提出有效的处理意见，主要表现在能根据事物的轻重缓急对工作的细节、策略、方法作出较为合理的规划。

**预测能力**——能提出较多有针对性的预测，对事件的观察力较好，有时考虑问题稍显片面，不能综合考虑各种影响因素，但对问题尚有一定的预测能力，对预测的实现充满信心。

**决策能力**——对复杂的问题能进行审慎的斟酌，能灵活地搜索各种解决问题的途径，并作出合理的评估对各种方案的结果做出清醒的判断，从而提出高质量的决策意见。

**沟通能力**——书面表达方面，语言较流畅，谈起问题来很有针对性，能提出有力的论据，结构性较强，表现出较为熟悉业务的各个领域，并能通盘考虑问题。

### 6. 本测验需要特别注意的事项

(1) 公文筐测验的适用对象为中、高级管理人员，它可以帮助企业选拔优秀的管理人才或考核现有管理人员。



(2) 公文筐测验从以下两个角度对管理人员进行测查：一为技能角度，主要考察管理者的计划、预测、决策和沟通能力；一为业务角度，公文筐的材料涉及财务、人事、行政、市场等多方面业务，它要求管理者具有对多方面的整体运作能力。包括对人、财、物流程的控制等。

(3) 公文筐测验对评分者的要求较高，它要求评分者了解测验的内核，通晓每份材料之间的内部联系，对每个可能的答案了如指掌，评分前要对评分者进行系统的培训，以保证测评结果的客观和公正。

## 第八章 基于情境的测验

### 第二节 无领导小组讨论

#### 一、什么是无领导小组讨论

##### (一) 无领导小组讨论的概念

1. 无领导小组讨论是指由一组应试者组成一个临时工作小组，讨论给定的问题、并作出决策。由于这个小组是临时拼凑的，并不指定谁是负责人，目的就在于考察应试者的表现，尤其是看谁会从中脱颖而出，成为自发的领导者。

2. 在无领导小组讨论中，不指定谁是领导，也并不指定每个应试者应该坐在哪个位置的，而是让所有应聘者自行安排、自行组织。

##### (二) 无领导小组讨论的优缺点

##### 1. 无领导小组讨论的优点：

- (1) 能检测出笔试和单一面试所不能检测出的能力或者素质；
- (2) 能观察到应试者之间的相互作用；
- (3) 能依据应试者的行为特征来对其进行更加全面、合理的评价；
- (4) 能够涉及到应试者的多种能力要素和个性特质；
- (5) 能使应试者在相对无意之中暴露自己各个方面的特点；
- (6) 能使应试者有平等的发挥机会从而很快地表现出个体上的差异；

(7) 能节省时间，并且能对竞争同一岗位的应试者的表现进行同时比较（横向对比）；

(8) 应用范围广，能应用于非技术领域、技术领域、管理领域和其他专业领域等。

## 2. 无领导小组讨论的缺点

(1) 对测试题目的要求较高；

(2) 对考官的评分技术要求较高，考官应该接受专门的培训；

(3) 对应试者的评价易受考官各个方面的影响（如偏见和误解），从而导致考官对应试者评价结果的不一致；

(4) 应试者有存在做戏、表演或者假装的可能性；

(5) 指定角色的随意性，可能导致应试者之间地位的不平等；

(6) 应试者的经验可以影响其能力的真正表现。

## （三）无领导小组讨论的取材

### 1. 讨论题的编制步骤

编制无领导小组讨论的试题通常有以下六个步骤：

(1) 工作调查：进行有关的工作分析，了解拟任岗位所需人员应该具备的特点、技能。根据岗位的这些特点和技能来进行有关试题的收集和编制。

(2) 案例收集：收集拟任岗位的相关案例，所收集的相关案例应该能充分地代表拟任岗位的特点，并且能够让应试者处理时有一定的难度。

(3) 案例筛选：对收集到的所有原始案例进行甄别、筛选，选出难度适中、内容合适、典型性和现实性均较好的案例。

(4) 编制讨论题：对所筛选出的案例进行加工和整理，使其符合无领导小组讨论的要求。主要包括剔除那些不宜公开讨论的部分。要设定一些与岗位工作相关而又符合讨论特点的情况或者问题，使其真正成为具备科学性、实用性、可评性、易评性等特点，成为既凝练又典型的讨论题。

(5) 讨论题的检验：讨论题编制完成以后，可以对相关的一组人（不是应试者）进行试测，来检查

该讨论题的优劣，检查讨论题能否达到预期的目的。

（6）讨论题的修正：检验完后，对于那些效果好的讨论题便可以直接使用，而对于那些不好的讨论题则要进行修正，直至其达到预期的效果。

## 2. 讨论题的要求

一般而言，对讨论的题目有以下五个方面的要求：

（1）讨论题的数量：对于每一组应试者而言，至少应有两个讨论题，以留作备用。在组与组之间，题目尽量不要相同，以免由于讨论题泄密造成评价效果上的不良影响。

（2）讨论题内容：所用的讨论题，在内容上应与拟任岗位相适应。

（3）讨论题难度：讨论题一定要一题多义，一题多解，有适当的难度，能够让讨论者争辩起来。

（4）讨论题立意：所用的讨论题。在立意方面，一定要高；讨论题的内容一定要具体，一定要避免那些玄妙、抽象、言之无物的争辩。

（5）角色平等：对于那些适用于角色分工的讨论题，讨论题本身对角色的分工在地位上一定要平等，而不能造成应试者之间有等级或者优劣的感觉。

## 3. 讨论题的形式

无领导小组讨论的讨论题一般都是智能性的题目，从形式上来分，可以分为以下五种：

（1）开放式问题：所谓开放式问题，是其答案的范围可以很广、很宽。

（2）两难问题：所谓两难问题，是让应试者在两种互有利弊的答案中选择其中的一种。

（3）多项选择问题：此类问题是让应试者在多种备选答案中选择其中有效的几种或对备选答案的重要性进行排序。

（4）操作性问题：所谓操作性问题，是给应试者一些材料、工具或者道具，让他们利用所给的这些材料，设计出一个或一些由考官指定的物体来，主要考察应试者的主动性、合作能力以及在实操任务中所充当的角色。

（5）资源争夺问题：此类问题适用于指定角色的无领导小组讨论。是让处于同等地位的应试者就有

限的资源进行分配，从而考察应试者的语言表达能力、分析问题能力、概括或总结能力，发言的积极性和反应的灵敏性等。

## 第八章 基于情境的测验

### 第二节 无领导小组讨论

#### 二、无领导小组讨论的应用

##### （一）无领导小组讨论的功能和特点

##### 1. 功能

通过模拟团队环境，检测应试者的组织协调、口头表达、洞察力、说服力、感染力、处理人际关系的技巧、非言语沟通（如面部表情、身体姿势、语调、语速和手势等）等各个方面的能力，以及自信程度、进取心、责任心、灵活性、情绪控制等个性特点和行为风格，以评价应试者之间的优劣。

##### 2. 特点

无领导小组讨论的突出特点是具有生动的人际互动性，能看到许多纸笔测验乃至面试看不到的现象，对预测真实团队中的行为有很高的效度。

##### 3. 经常考察的要素

以下是无领导小组讨论的维度定义：

（1）组织行为。

（2）洞察力。

（3）倾听。

（4）说服力。

（5）感染力。

（6）团队意识。

（7）成熟度。

## （二）无领导小组讨论的应用

### 1. 适用对象

无领导小组讨论常用于选拔管理人员，它的适用对象为具有领导潜质的人或某些特殊类型的人群（如营销人员），可以从中择优选拔企业所需的优秀人才。

### 2. 测验时间

无领导小组讨论的时间一般是 30—60 分钟左右。整个讨论可分为 3 个阶段。

### 3. 测验的施测过程

### 4. 测验的样题（含指导语）

指导语：

**现在我们要根据企业的要求开一个讨论会。在座的各位现在就组成一个专题工作小组。现在公司要对下列问题进行讨论、分析，并作出决定。**  
**请大家充分讨论，并拿出小组的意见来。讨论共有 35 分钟。请大家充分利用时间。**  
**讨论一旦开始，将不再回答你们的任何提问，也不干预你们的讨论。**

样题：

**你认为什么样的领导是好领导？以工作为取向还是以人为取向？**

### 5. 报告样例

#### 无领导小组讨论结果

**考号：165 姓名：\*\*\* 性别：男 年龄：25**

1	2	3	4	5	6	7
组织行为	洞察力	倾听技巧	说服力	感染力	团队意识	成熟度
88.6	74.8	86.4	54.4	71.8	88.3	68.2

评语：

**具有良好的领导组织意识，善于影响他人，营造合作的团队氛围，表现出较强的组织能力，**

**对问题的思考和分析有良好的洞察力，能比较深入地把握关键的事物，做出客观合理的反应，**

**在与人合作中有较好的沟通交流能力，团队合作意识强，人际关系取向较高，**

**在言谈、举止上成熟度一般，说服力有待加强。**

#### 6. 本测验需要特别注意的事项

（1）评分者在观察被试的行为和言语表现时，不要因为被试的某些人格特点而对被试造成不应有的偏见，这样会使结果失之偏颇。

（2）评分过程中，要求多名评分者对同一被试的不同能力要素分别打分，取其平均值作为被试的最后得分，这样的结果才科学、公正。

### 第八章 基于情境的测验

#### 第三节 面试

##### 一、什么是面试

###### （一）面试的定义

面试是在考官与应试者直接交谈或设置应试者于某种特定情境中进行观察，了解应试者素质状况、能力与个性特征及求职动机等情况，从而完成对应试者的评价的一种十分有用的测评技术。

###### （二）面试的特点

###### 1. 重过程

###### 2. 综合性

###### 3. 灵活性

###### 4. 行为导向

###### 5. 过分依赖主试

### （三）面试的理论依据

1. 行为反应了人的心理特点

2. 社会心理学的研究

（1）言语行为（7%）

（2）非言语行为（93%）

A. 动态无声的身体语言：眼神、表情动作、身体运动

B. 静态无声的身体语言：身体姿势、空间距离等

C. 副言语：重音、抑扬顿挫等

3. 面试中的无意识流露——精神分析的观点

## 第八章 基于情境的测验

### 第三节 面试

#### 二、面试的分类

##### （一）情景面试

1. 面试题目是假定的情景
2. 通过求职者对情景的反应作出评价
3. 情景试题来源于工作或者能体现出工作所需的素质
4. 对事例的反应能区分开优秀员工和一般员工

##### （二）基于行为的面试

1. 用过去预测未来表现
2. 定义素质的典型行为（正/负）
3. 收集应聘者过去的 STAR：Situation/Target/Action/Results

##### （三）小组面试

1. 多个面试考官
2. 每个考官问问题
3. 考察全面
4. 给求职者增加额外压力

#### （四）压力面试

1. 考察求职者如何承受压力
2. 问题使求职者感到压力或者陷入尴尬

#### （五）非结构化面试

1. 开放式任意式谈话
2. 没有固定模式
3. 没有事先准备好的问题
4. 求职者有较大自由度决定讨论的方向
5. 灵活性强，可以深入
6. 对主考官要求较高

#### （六）结构化面试

与一般面试相比，结构化面试对面试的考察要素、面试题目、评分标准、具体操作步骤等进一步规范、结构化和精细化，并且统一培训面试考官，提高评价的公平性，从而使面试结果更为客观、可靠，使同一个职位的不同应试者评估结果之间具有可比性。

##### 1. 结构化面试的结构化表现

- （1）面试试题（内容、种类、编制）结构化：问题基于考核要素
- （2）评分标准结构化（测评指标、水平刻度、测评规则）
- （3）面试考官结构化
- （4）面试过程结构化



## 2. 结构化面试的程序

(1) 工作分析，确定测评要素以及权重。

(2) 制定测验题目和答案。

(3) 对考官进行培训，明确测评要素的具体含义，掌握评分标准（讲解示范、模拟演练、观察评价）。

(4) 组织实施面试。

### (七) 半结构化面试

介于结构化和非结构化面试之间。

## 第八章 基于情境的测验

### 第三节 面试

#### 三、面试题目的结构与类型

##### (一) 面试题目的结构

1. 题干：文字要精炼，表意明确，措辞严谨。

2. 追问：根据题干的外延和内涵进行设计。

3. 测评要素：每道题目要明确测评要素，并非一个题目考察一个要素。但是，在面试之前要认真提炼。

4. 评分要点：帮助主试把握评判应试者素质的主要着眼点和行为。

评分要点包含两个部分：该题目的设计思路；指出对应试者应该观察和评价的若干方面，以及相关行为与结果标准。

##### (二) 面试题目的类型

###### 1. 背景性问题

(1) 营造良好的沟通气氛，缓和被试的紧张情绪。

(2) 对被试有基本的了解，为进一步交流提供有价值的话题。

## 2. 智能性问题

(1) 考察被试的综合分析能力、逻辑思维能力、言语表达能力。

(2) 并非单纯的智力问题，而是一些值得思考和争论的各种现实问题。

## 3. 工作知识问题

与工作岗位的知识有关。

## 4. 情境性问题

(1) 描述一个针对相关能力的、与工作有关的假定情境，要求被试回答在这个给定的情境中他们会怎么做。

(2) 类似于情境模拟，是一个语言模拟的题目。考察多种能力，包括决策能力、思维敏捷性、随机应变能力。

## 5. 行为性问题

(1) 让被试讲述过去的某个经历，从而获得被试的能力要素的信息。

(2) 讲述中遵循 STAR 原则。

## 6. 意愿性问题

(1) 考察被试的价值取向、报考动机与职位要求的匹配性，以及生活态度等个性倾向。

(2) 可以直接提问，也可以采用投射和迫选技术。

# 第八章 基于情境的测验

## 第三节 面试

### 四、主试与被试应该注意的问题

#### (一) 主试应该注意的问题

##### 1. 主试常犯的错误

###### (1) 第一印象

(2) 近因效应

(3) 对比效应

(4) 刻板效应

(5) 晕轮效应

(6) 类我效应

## 2. 面试中的观察

(1) 坚持观察的综合性、目的性和客观性原则

(2) 避免以貌取人或者光环效应

(3) 注意面部表情

(4) 注意身体语言

(二) 被试在面试中的技巧

(1) 印象管理的技术

(2) 分析职位要求

(3) 分析主试的相关信息

(4) 适度紧张、自信、平和、有一定的亲和力

## 第八章 基于情境的测验

### 第四节 评价中心

#### 一、什么是评价中心

##### (一) 评价中心的定义

所谓评价中心，是指在有限的时间和特定的环境中，运用多种方法和技术，同时对多个人进行管理素质测量和评价的一套标准化程序。主要测评手段包括无领导小组讨论、角色扮演、心理测验、公文筐测验、演讲、案例分析、搜寻事实、面试。

评价中心与其他测量工具之间的关系。

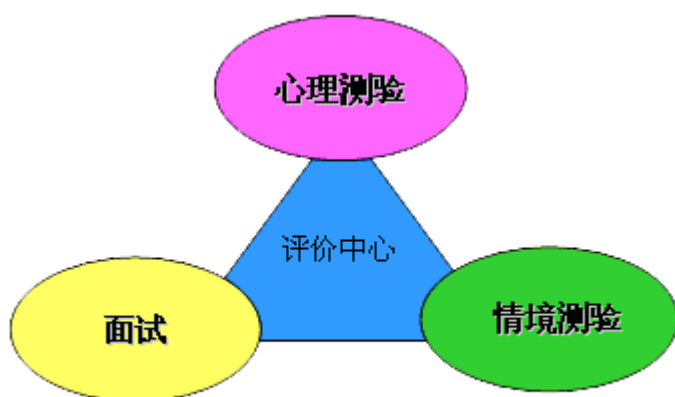


图8-1 评价中心与不同测量工具之间的关系

## （二）评价中心的特点与作用

### 1. 评价中心的特点

- （1）情景模拟原理：测评内容的模拟化
- （2）综合性：信息多元化、多种方法的组合
- （3）灵活性：手段多样化
- （4）互动性：测评过程的动态化和交互作用
- （5）测评目标与工作的关联性
- （6）评价者培训
- （7）成本高

### 2. 评价中心的作用

随意选拔，15%；经理提名，35%；评价中心，70%。

## 第八章 基于情境的测验

### 第四节 评价中心

#### 二、评价中心其他方法简介

##### （一）角色扮演

- 1. 评价者或者评价者助手扮演角色，经过培训，形成一个标准化模式。
- 2. 扮演什么样的角色决定于考察要素。
- 3. 考察说服能力、表达能力、处理冲突的能力、思维灵活性和敏捷性等。

4. 对扮演者要求较高。

#### （二）演讲

1. 分为竞选演说、辩论式演讲、陈述观点演讲。
2. 一般结合其他测评方法（如公文筐测验）。
3. 考察分析推理能力、语言表达能力以及在压力下的反应。

#### （三）搜寻事实

##### 1. 测验的基本过程

- （1）参加者被提供一个与某一问题的相关（模糊）信息；
- （2）要求参加者从相关人员那里获得信息，然后在一定的时间限制内做出决策；
- （3）参加者作一个简短的汇报，信息被收集人作一个评价。

##### 2. 考察的能力

- （1）智力（思维清晰、敏锐度，思维的条理性和抽象能力）；
- （2）实际判断能力（修正结论的能力、企业的一般常识等）；
- （3）获得信息的能力；
- （4）分析理解问题能力；
- （5）社会知觉；
- （6）决策能力。

##### 3. 注意事项：编制者要准备好充足的信息源。

#### （四）案例分析

##### 1. 基本过程

- （1）提供有关组织中问题的材料。
- （2）要求求职者给出解决建议。

##### 2. 考察的能力：综合分析能力和判断决策能力。

##### 3. 主要问题：评分困难。



职业经理 MBA 整套实战教程

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第九章 测量的信度

### 第一节 测量的三种误差

#### 一、什么是误差

##### （一）概念

误差是在测量中与目的无关的变量（原因）所产生的不准确或不一致效应（准确性与一致性的含义见图9-1）。

任何测量都不是绝对准确的，不准确的数据就有误差。

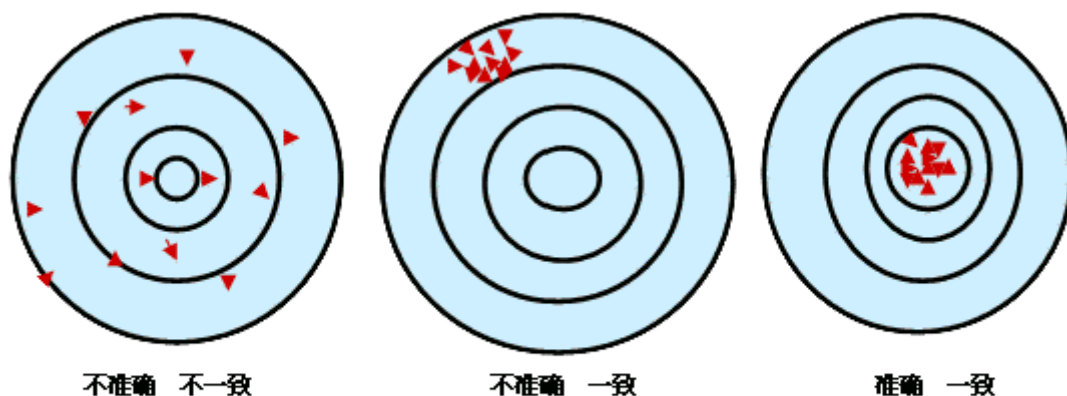


图9-1 准确性与一致性的含义

## （二）含义

1. 误差是由与测量目的无关的变量引起。
2. 误差会导致不准确或不一致的结果。

## （三）来源

1. 测验内部（测量工具本身）引起的误差，例如测量工具本身不能测量所要测的能力。
2. 施测过程引起的误差，例如测量过程中指导语不清楚，环境对测量的影响等。
3. 被试本身引起的误差，例如被试的焦虑情绪、参加测量的动机、对测量的练习因素和其本身具备

的应对测验的技巧，以及被试的反应定势（所谓反应定势是指独立于测验内容的反应倾向）。常见的反应定势有以下几种：

- （1）求“快”和求“精确”的反应定势
- （2）偏好正面描述的反应定势
- （3）偏好特殊位置的反应定势
- （4）偏好较长选项的反应定势
- （5）猜测的反应定势

4. 主试本身引起的误差，例如主试的态度、动机、人格、期望（罗森塔尔效应）等都会对测量结果产生影响。

5. 主试、被试交互作用所引起的误差。

## 二、误差的分类

### （一）随机误差

随机误差是一种可变误差，它由与测量目的无关的偶然因素引起，不易控制，使多次测量产生了不一致的结果，这种误差是难免的，但却是毫无规律的。例如：称重量的时候，不同人对秤杆高度的掌握不同。

### （二）系统误差

系统误差是由与测量无关的变量引起的一种恒定而有规律的误差，稳定地存在于每一次测量中，测值一致，但不准。例如在秤砣上搞鬼，就形成了系统误差。

### （三）过失误差

不正确使用测量工具、记录错误、仪器失灵等失误造成的误差称为过失误差，因此在收集资料时，应该避免过失误差，处理数据时，也必须把包含过失误差的数据去除。

## 三、误差的影响

（一）系统误差只影响测值的准确性，与效度有关，不影响信度。

（二）随机误差既影响准确性，又影响一致性，与效度和信度都有关。

## 第九章 测量的信度

### 第二节 信度的含义与作用

#### 一、信度的含义

##### （一）真分数的定义

1. 真分数的概念定义：一个测量工具在测量没有误差时所得到的值。

2. 真分数的操作定义：经过无数次测量所得到的平均值。

（二）信度主要是指测量结果的可靠性或一致性。由于接受测量时应试者的行为可能会由于各种原因而产生变动，偏离了其真实行为，这就会使测量结果产生误差。测量结果的可靠性与测量结果受误差影响的程度密切相关。误差大，分数的可靠性就降低。信度便是说明测量的可靠性或一致性的指标。

（三）在研究信度时，主要应考虑两方面的问题：

1. 稳定性问题，涉及：

- （1）测验分数一致性的程度，即不同时间、不同测验条件下所得分数之间的一致性有多大；
- （2）一个人的获得分数与“真实分数”之间接近程度如何；
- （3）测验的测量一致性是否可以达到实际应用的程度等问题。

2. 影响稳定性的原因，即分数不稳定、不一致的原因：

- （1）什么因素造成了这种差异；
- （2）这些效应的相对作用如何。

## 二、信度的作用

信度高低的指标通常以相关系数表示，称为信度系数（reliability co-efficient）。信度系数一般是同一样本所得的两组资料的相关，在理论上表示为实得分数与真实分数相关的平方：

$$r_{XX} = r_{XT}^2 = S_T^2 / S_X^2$$

其中， $r_{XT} = S_T / S_X$  有时也称为信度指数，它是真实分数标准差与实得分数标准差的比率。

另外，还可通过分析个人分数再测时的变化（误差）来考察信度。我们知道两次测验的分数不可能完全一致，但是误差的大小与测验的信度有直接关系：两次测验分数的差异越大，信度就越低。在人事测量中，对信度系数进行确定通常有以下两方面的作用：

（一）解释真实分数与实得分数的相关

信度可以解释为总的方差（即标准差的平方的缩略语）中有多少比例是由真实分数的方差决定的，即测量分数的变化中有多少是真正反映了被测者分数的变化。



（二）说明可以接受的信度水准

信度究竟要多高才是可以接受的呢?一般说来，最理想情况为 1.00，但实际上是达不到的。不过我们可以用已有的同类测验作为比较的标准。表 9—1 表示出几种类型的测量工具的信度系数。一般的能力与成就测验的信度系数在 0.90 以上；人格、兴趣等测验的信度系数通常在 0.80—0.85 之间。

测验类型	信度		
	低	中	高
成就成就测验	0.94	0.92	0.98
学术能力测验	0.94	0.90	0.97
成在倾向性测验	0.26	0.88	0.96
客观人格测验	0.46	0.85	0.97
兴趣问卷	0.42	0.84	0.93
态度量表	0.47	0.79	0.98

一般来说，当  $r_{xx} < 0.70$  时，不能用测验对个人作评价，也不能在团体间作比较；当  $r_{xx} > 0.70$  时，可用于团体间比较； $r_{xx} > 0.85$  时，可用于鉴别个人情况。

第九章 测量的信度

第三节 信度的分类

一、重测信度与复本信度

（一）重测信度

1. 重测信度的概念

（1）重测信度（test-retest reliability）又称为稳定性系数，它的计量方法是采用重测法：用同一测验，在不同时间对同一群体施测两次，这两次测量分数的相关系数（采用积差相关系数）即为重测系数。根据重测相关系数的高低，可以得知测量结果在经过一段时间之后的稳定程度。重测信度越高，说明测量的结果越一致、越可靠。人事测量所测查的特质中，有相当一些内容都具有一定的稳定性，如人格、基本能力倾向等，因此，对于这些内容来说，测量工具的信度是十分重要的。

(2) 重测信度所考察的误差来源是时间的变化所带来的随机影响，例如由气候、偶然的噪音或其他干扰，以及引起被试本身身心状态变化的因素如疾病、疲劳、情绪波动、焦虑等原因造成的对测量结果的影响。重测信度代表测量能应用于不同时间的程度，信度越高，受环境中日常随机因素的影响越小，测量结果越稳定。

## 2. 重测信度评估应注意的问题

(1) 在评估重测信度时，必须注意重测间隔的时间。重测间隔时间的长短，必须根据测验的性质和目的来确定。如果希望测量结果能预测较长时间的变化，则重测间隔的时间应该长一些。例如，对于人格测验，重测间隔在两周到六个月之间比较合适。

(2) 重测信度一般只反映由随机因素导致的变化，而不反映被试行为的长久变化。例如，应试者智力的发展与能力的提高，不是重测信度考虑的因素。由于这些因素导致的重测相关系数的降低，不能说明测验的重测信度低。

(3) 不同的行为受随机误差影响不同。

## 3. 重测信度的前提假设

重测信度的前提假设是每个应试者对前一次测验的遗忘程度相同，而且在重测间隔期间没有学过与测量有关的其他材料，或者说每人所学习的程度相同。这个假设在现实生活中不大容易完全满足。另外，有些解决问题型测验不宜采用重测信度，因为应试者一旦知道答案就不容易忘记，从而造成假性高相关。只有不大容易受重复影响的测验，如感觉—运动测验或人格测验，比较适合用重测法计算信度系数。

## (二) 复本信度

### 1. 复本信度的概念

(1) 复本信度 (alternative-form reliability) 又称等值性系数，它是以两个测验复本 (功能等值但题目内容不同) 来测量同一群体，然后求得应试者在这两个测验上得分的相关系数 (积差相关)。复本信度的高低反映了这两个测验复本在内容上的等值性程度。两个等值的测验互为复本。

(2) 计算复本信度的主要目的在于考察两个测验复本的题目取样或内容取样是否等值。例如，同样

是测量数学运算能力的测验，如果一个测验复本侧重于考核加减法的运算，而另一个复本侧重乘除法的运算，两者之间的相关必定不会太高，即复本信度低。测验的复本信度低，说明必定有一项测验复本的取样有问题。复本信度考虑的就是这种内容取样误差的影响问题。

(3) 复本信度也考虑两个复本实施的时间间隔。如果两个复本几乎是在同一时间实施的，相关系数所反映的才是不同复本的关系，而不掺有时间的影响。在实际工作中，为了避免施测顺序效应，常常是一半被试先做其中一个复本，而另一半被试先做另一个复本。

## 2. 复本信度的主要优点

- (1) 能够避免重测信度的一些问题，如记忆效果、练习效应等；
- (2) 适用于进行长期追踪研究或调查某些干涉变量对测验成绩的影响；
- (3) 减少了辅导或作弊的可能性。

## 3. 复本信度的局限性

- (1) 如果测量的行为易受练习的影响，则复本信度只能减少而不能消除这种影响；
- (2) 有些测验的性质会由于重复而发生改变，例如某些问题解决型的测验，如果掌握了解题原则，就有可能产生迁移，尤其当复本只是在题目具体内容上有改变时，这种正迁移的作用会很强；
- (3) 有些测验很难找到合适的复本。

# 第九章 测量的信度

## 第三节 信度的分类

### 二、内部一致性信度和评分者信度

#### (一) 内部一致性信度

内部一致性信度 (consistency reliability) 系数主要反映的是测验内部题目之间的关系，考察测验的各个题目是否测量了相同的内容或特质。内部一致性信度又分为分半信度和同质性信度。

#### 1. 分半信度

### （1）分半信度的含义

分半信度（split-half reliability）系数是通过将测验分成两半，计算这两半测验之间的相关性而获得的信度系数。

### （2）分半信度的计算

①采用分半法计算分半信度，首先要解决的问题是如何将测验分成可比较的两部分。

②计算分半信度系数仍可以采用常用的积差相关方法。但是，这种相关系数实际上只是半个测验的相关系数。常用的修正公式是斯皮尔曼-布朗公式。分半法中的斯皮尔曼-布朗修正公式为：

$$r_{xx} = 2r_{hh} / (1 + r_{hh})$$

其中为  $r_{xx}$  为两半测验的相关系数， $r_{xx}$  为估计或修正后的信度。

当两半测验的方差不同时，应采用卢伦公式或弗朗那根公式进行修正。卢伦公式只要求将被试在两半测验的分数之差的方差（ $S_d^2$ ）和测验总分的方差（ $S_x^2$ ）代入如下公式即可直接计算分半信度：

$$r_{xx} = 1 - \frac{S_d^2}{S_x^2}$$

弗拉那根公式也可直接计算分半信度：

$$r_{xx} = 2 \left( 1 - \frac{S_a^2 + S_b^2}{S_x^2} \right)$$

其中， $S_a^2$  与  $S_b^2$  为两个分测验分数的方差； $S_x^2$  为总分方差。

## 2. 同质性信度

### （1）同质性信度的含义

①同质性是指所有测验题目测量的只是单一特质或内容，表现为所有测验题目得分的一致性。例如，“1+1=?”和“2+2=?”这两个简单加法题可以说是高度同质性的。

②同质性是保证测验只测量单一特质的必要条件。如果同质性差，则测验可能混淆了不同的内容，其

结果就无从判断究竟反映了被测者的什么特征。

③同质性信度（homogeneity reliability）是指测验内部的各题目在多大程度上考察了同一内容。

④所有题目都应该保证只测量一种特质或内容。如果需要在同一测验中测量不同的内容，就应该将测验设计为几个分测验，每个分测验测量一种内容。

## （2）同质性信度的计算

常用的同质性信度计算方法是库德—理查逊估计方法。计算公式有库德—理查逊 20 号公式（简称 KR20）和 21 号公式（KR21）。下面是 KR20 公式：

$$r_{\text{KR20}} = \frac{n}{n-1} \left[ 1 - \frac{\sum_{i=1}^n p_i q_i}{S_x^2} \right]$$

其中， $n$  是题目数， $p_i$  为答对第  $i$  题的人数的比例， $q_i$  为答错第  $i$  题的人数的比例， $S_x^2$  为测验总分的变异。此公式是由库德和理查逊于 1937 年提出的，仅适用于（0、1）记分的测验。

由于在进行项目分析时，要求算出各个题目的难度  $P$ ，因此使用该公式非常简便。在各试题难度相同或接近的情况下，还可以采用计算更为简便的 21 号公式，但当试题难度相差悬殊时，采用 KR21 号公式有低估信度系数的可能。

$$r_{\text{KR21}} = \left[ \frac{K}{K-1} \right] \left[ 1 - \frac{k \bar{P} \bar{Q}}{S_x^2} \right]$$

其中，各指标含义与 KR20 相同，只是  $\bar{P}$  与  $\bar{Q}$  分别表示题目的平均通过率和失败率。此公式只有当所有题目的难度接近时才适用。

## 3. 分半信度与同质性信度的差别

同质性信度和分半信度虽然都是内部一致性信度，但两者之间存在差别。例如，假设有一项 50 题的测验，有 25 种题目，如第 1 题和第 2 题为词汇题，第 3 题和第 4 题为算术题，第 5 题和第 6 题为空间推理题，等等，如果分成奇偶两半，其得分的相关系数必然很高，但其同质性实际上是很低的，因为被试在

这 50 道题上的得分一致性不会很高。显而易见，使用库德—理查逊方法估计的信度要低于分半信度，后者可作为测验同质性的粗略指标。

库德—理查逊方法适用于二分法计分的项目，但有许多测验项目采用多重计分，例如人格测验、态度量表等，被试的反应常常分为“经常”、“有时”、“偶尔”、“从不”等各种等级，分数也因而是多级的。这种类型的测验，经常采用克伦巴赫  $\alpha$  系数来表示内部一致性的高低：

$$\alpha = \left[ \frac{n}{n-1} \right] \left[ 1 - \frac{\sum_{i=1}^n S_i^2}{S_x^2} \right]$$

其中， $S_i^2$  表示所有被试在第  $i$  题上的分数变异，其余指标的含义与  $KR_{20}$  相同。此公式是由克伦巴赫提出的，它不要求测验题目仅是 (0、1) 记分，可以处理任何测验的内部一致性系数的计算问题。

**例如，某态度量表共 5 题，被试在各题上得分的方差分别为 0.80、0.81、0.79、0.78、0.82，测验总分的方差为 15.00，因此测验的  $\alpha$  系数为：**

$$\alpha = \left[ \frac{5}{5-1} \right] \left[ 1 - (0.80+0.81+0.79+0.78+0.82)/15 \right] = 0.92$$

## (二) 评分者信度

在有些测量情形中，评分者的评判也是误差的来源之一。例如投射测验、创造力测验、无领导小组讨论、管理者情景模拟测验等，都依赖于评分者的判断，这种判断的主观性往往造成不同评分者的评分很不一致，因此也有必要考虑评分者信度。

评分者信度 (raters reliability) 是指不同评分者对同样对象进行评定时的一致性。当由两个独立的评分者打分时，一般可以采用积差相关方法，也可以采用斯皮尔曼等级相关方法。如果评分者在三人以上，而且又采用等级计时，就需要用肯德尔和谐系数来求评分者信度。其公式为：

$$W = 12 \cdot \frac{\sum_{i=1}^N R_i^2 - \frac{(\sum_{i=1}^N R_i)^2}{N}}{K^2(N^3 - N)}$$

其中， $K$  是评分者人数， $N$  是被评的对象数 (通常是考生数，每个考生一份试卷)， $R_i$  是第  $i$  个被评

对象（考卷）被评的水平等级之和。

例如，有 4 名评分者，对 6 份答案进行评分，所评等级如下：

评分者	答卷编号					
	一	二	三	四	五	六
甲	4	3	1	2	3	6
乙	3	3	2	1	4	6
丙	4	1	2	2	3	6
丁	6	4	1	3	3	3
$\Sigma$	19	11	6	8	17	23

可求得：

$$\Sigma R_j = 19 + 11 + 6 + 8 + 17 + 23 = 84$$

$$\Sigma R_j^2 = 19^2 + 11^2 + 6^2 + 8^2 + 17^2 + 23^2 = 1400$$

$$S = 1400 - 84^2 / 6 = 224$$

$$R^2 = 12 \times 224 / [4^3 \times (6^3 - 6)] = 0.80$$

第九章 测量的信度

第四节 信度的影响因素

一、样本的性质

（一）样本团体的分数分布

任何以相关系数表示的信度系数都会受样本团体分数分布的影响。分数分布越广，信度系数就相对越高，分数分布越窄，信度系数就会越低。例如，假设有 A、B 两项测验，其中 A 测验的样本团体的分数分布为 20—90 分，而 B 测验的样本团体的分数分布在 80—90 分之间。显然，在第二次重测时，相同的分数起伏对 A、B 这两项测验的影响不同。在测验 A 中，被试的成绩波动 1、2 分，分数的相对位置的变化不大。而在测验 B 中，被试的成绩波动 1、2 分就可能导致其分数的相对位置发生很大的变化。因此，测验 A

的重测信度必然高于测验 B。

## （二）样本团体的异质性

信度系数还会受到样本团体异质性的影响。一般来说，取样团体的异质性越大，信度系数就相对越高。例如，有一项测验，分别施测于 A、B 两组被试。其中 A 组被试较为异质（如各年级的学生都有），其分数分布为 20—90 分，而 B 组被试同质程度较高（都为二年级学生），其分数分布为 80—90 分。显然，由于 A 组被试的分数分布比 B 组要广得多，按照上面所提到的分数分布与信度的关系，我们可以推知，测验以 A 组被试为样本团体得到的信度，要比采用 B 组被试得到的信度高。所以说，同质性样本团体的分数分布窄，得到的信度低，而异质性样本团体的分数分布得到的信度较高。

## （三）不同团体间能力水平的差异

测验所施测的团体的平均能力水平的不同也会对信度产生影响。例如，在斯坦福一比奈量表中，不同年龄组的信度从 0.83—0.98 不等。因为对于年幼的团体，他们的平均能力水平低，他们的分数基本上是凭猜测获得的，其靠猜测的测验结果总是不会很稳定的，所以信度值较低。这种情况导致的信度偏差，很难用一般的统计公式来校正，只能通过对各种年龄及能力水平的检验来确定。

# 第九章 测量的信度

## 第四节 信度的影响因素

### 二、测验的长度与测验的难度

#### （一）测验的长度

信度还会受测验长度（即题目的多少）的影响。一般来说，测验越长，信度值越高。一方面，测验越长，题目取样或内容取样就越充分，结果就越可靠。举一个极端的例子，如果词汇量的测验中只包含一个试题，仅依据应试者对一个单词的记忆，来确定应试者的词汇量的大小，其结果必定是非常不可靠的。另一方面，较长的测验也不容易受到猜测的影响。猜对一个题目的概率很高，但如果测验足够长，猜对所有题目的概率就微乎其微了。



通过下面的斯皮尔曼—布朗公式，可以估计增长或缩短一个测验对其信度系数的影响有多大：

$$r_{nn} = \frac{nr_u}{1 + (n-1)r_u}$$

其中， $r_{nn}$ 为估计或修正后的信度； $r_u$ 为实得的相关系数； $n$ 为测验增长或缩短的倍数。

在增加测验长度的时候要注意：只有所增加的题目和原题目在性质上相同时，才能达到提高信度的效果。

## （二）测验的难度

测验的难度也会对信度产生影响。如果一个测验的难度太低，测验分数会非常集中并聚在高分端，即出现天花板效应，或者难度太大，所有分数都集中在低分端，出现地板效应，都会使测量到的分数分布太窄，导致信度降低。

只有当测验的难度水平能够使测验分数分布范围最大时，测验的信度才会比较理想。一般来说，当所有被试的平均分数从零分到满分分布时，测量的信度最高。因此，测验的长度和难度会共同起作用。如果只增加测验，测验分数不能充分散开，那么增加长度的努力也会变成徒劳。

# 第十章 测量的效度

## 第一节 效度的含义

### 一、效度的理论定义

从测量理论角度讲，效度可以定义为：与测量目标有关的真实分数方差与总分方差的比率。“有关的真实分数方差”是由测验所要测量的目标变量所产生的方差。效度的定义用公式可以表示为：

$$r_{XY} = S_Y^2 / S_X^2$$

其中,  $r_{XY}$  为测验效度;  $S_V^2$  为有效方差;  $S_X^2$  为总方差。

效度定义也可以通过真分数的概念和方差(变异数)分析方法加以说明。我们知道, 一组测验分数的总方差等于真实方差与误差方差之和:

$$S_X^2 = S_T^2 + S_E^2$$

而真实方差还可分为两部分:

$$S_T^2 = S_V^2 + S_I^2$$

其中,  $S_V^2$  代表有关(有效)方差;  $S_I^2$  代表无关而稳定的方差。将以上公式代入上面的公式可得:

$$S_X^2 = S_V^2 + S_I^2 + S_E^2$$

## 二、信度和效度的关系

根据定义, 我们知道信度和效度的差别在于所考虑的误差不同。信度考虑的是随机误差的影响; 效度的误差则还包括对测验目的来说无关的变量所引起的系统误差。

由上述公式不难发现, 当测量工具的误差方差( $S_E^2$ )减小时(亦即测量更稳定可靠时), 潜在的有效方差( $S_V^2$ )比例有可能增加, 但是其余的系统误差方差可能大, 也可能小, 所以说误差方差低(即信度高)并不保证效度就一定高; 但效度高的话, 信度必然要高。信度是效度的必要条件。

## 三、效度的性质

(一) 需要注意的是, “测验的效度”是简便的说法, 实际上效度指的是“测量结果的效度”, 是指测量结果的准确性程度, 而并不是指测验本身。

(二) 效度是连续性的, 效度高低只是程度上的差别, 它不是“全有”或“全无”的变量。因此, 我们应避免评价某一测验的结果为“有效”或“无效”, 而应区分为效度较高或较低。

(三) 效度同时也是针对测量目标而言的, 反映了测量结果对测量目标的体现程度。因此, 在评价和

使用某项测验时，应该根据所要使用的用途，有针对性地考察其效度。

## 第十章 测量的效度

### 第二节 效度的分类

#### 一、内容效度

##### （一）内容效度的定义

1. 内容效度（content validity）是检查测验内容是否是所欲测量的行为领域的代表性取样的指标。例如，成就测验往往采用内容效度的评估方法。

在实际工作中，我们编制的测验不可能包含所要测量的行为领域的全部可能的材料或情境，只能选择一个有代表性的样本，通过观察被试对个别题目的反应，来推测他的总体行为的表现。因此，取样的恰当性就是影响测量效果的一个重要因素。

##### 2. 较好的内容效度依赖于两个条件

（1）测验内容范围明确。

（2）测验内容的取样有代表性。

##### 3. 表面效度与内容效度的关系

表面效度并非真正的测验效度，而是外行人对测量对象的臆断，可用来操作施测对象的测验动机。但并不具有专家的判断水准。对于最高行为测量，表面效度越高越好。对于典型行为测量，表面效度越低越好。

##### （二）内容效度的评估方法

##### 1. 专家判断法

确定内容效度的方法通常是由专家根据测验题目和假设的内容范围作系统的比较判断。如果专家们认为测验题目恰当地代表了所测内容，则测验具有内容效度。这种方法的主要问题是缺乏一种数量化指标来描述内容效度的高低；不同判断者的判断可能不一致；如果测验内容范围缺乏明确性，会使效度的判断过

程发生困难。为了使内容效度的确定过程更为客观，可以依次采用如下步骤：

- （1）确定总体范围，即描述有关的知识与技能及所用材料的来源；
- （2）编制双向细目表，确定内容和技能各自所占的比例，并由测验编制者确定各题所测的是何种内容与技能；

表 10-1 成就测验的“测量工具的编制与实施”课程内容的测量工具细目表

行为目标	内容（材料）				
	测量工具准备 （10%）	测量工具的编制 （20%）	实施 （14%）	记分 （22%）	项目分析 （22%）
概念知识	工作分析 效标事件 显示性测量工具	匹配题 反应定势	关系效应 晕轮效应	记分键 组合记分 机器记分	效标 内部一致性 测量工具同质性
具体事实 知识	教育目标分类的 类别	问答题和客观题 的优缺点	测量工具成就的 影响因素	问答记分和客观 记分的规则	决定项目效度的 方法 项目分析的目的
理解	解释编制测量工 具计划的目的	（0 题）	标准化	记分起加权的作用	PD 关系的理解
应用	编制测量工具领 域某一个单元的 双向细目表	测量理解、应用、 分析综合、评价 的多选題实例	编制测量工具附 示语	项目的修正加权	速度与区分度的 计算
总数	8 题	13 题	7 题	11 题	11 题

- （3）制定评定量表来测量测验的整个效度及其他特点，如测验包括的内容、技能、材料的重要程度、题目对内容的适用性等。由每位评判者在评定量表上作出判断，总结获得测验内容效度的证据。

2. 统计分析的方法

- （1）用复本信度来估算内容效度：克伦巴赫提出，内容效度可以由两个独立的但取自同样内容范围的测验得分的相关（即复本信度）来作数量上的估计。若两者相关高，就有证据表明它们具有内容效度；若相关低，这两者中至少有一者缺乏内容效度。

- （2）评分者信度：计算两个评分者之间评定的一致性，即考虑评分者信度。虽然它所代表的是判断信度，但由于来自两个独立的评判者，因此符合程度越高越能反映测验的内容效度。

3. 再测法和经验法

先施测于某个团体，得出一个分数。然后进行培训与练习。如果成绩提高了，则证明测验是培训的内容。具有较高的内容效度。

内容效度的确定还可采用经验的方法，例如，对于成就测验可以检查不同年级的被试的总分和每题分数变化的情况。一般而言，如果随着年级增高，被试的分数和每个题目的通过率也随着升高，就可以推测该测验基本测量了学校的教学内容和目标，这也是测验的内容效度的证据之一。

### （三）内容效度的应用

1. 内容效度最适合于评估教育和职业成就测验。在这种应用中，通过对内容效度的评价可以回答以下两个问题：（1）该测验是否适应考察的某种技能和知识的代表性样本；（2）测验的成绩是否不受无关因素的影响。

2. 内容效度也适合于某些用于选拔和分类的人事测验。这种测验中，测验内容是实际工作的一个样本，应包含实际工作所需要的技能和知识。在这种情况下，应该通过内容效度的分析来确定测验是否的确测量了实际工作中所需要的知识和技能。

3. 内容效度的评估一般不适用于能力倾向测验和人格测验。

## 第十章 测量的效度

### 第二节 效度的分类

#### 二、效标关联效度

##### （一）效标关联效度（criterion—related validity）的含义

效标关联效度也称效标效度（criterion validity），反映的是测验分数与外在标准（效标）的相关程度，即测验分数对个体的效标行为表现进行预测的有效性程度。

##### （二）效标和效标测量

1. 效标关联是考察测验效用的外在参照标准。例如一个机械能力倾向测验，其效度的标准（效标）可以是成为机械师后的工作表现；对于一个管理能力测验而言，其效标可以是将来管理工作的绩效。效标

关联效度往往用于预测性测验。这种测验中，根据测验分数作出的预测一般用于甄选决策，所以，只有当证明测验分数确实能够预测所欲研究的行为时，这种决策才可能正确。

效标是衡量测验有效性的参照标准，是一种可以直接、独立测量的行为。换句话说，要测量效标，就必须把效标行为转化为某种可以操作的测量指标，以便进行比较。这种可操作的测量指标就称为效标测量。因此，效标的概念可以细分为观念效标（conceptual criterion，即效标的实质概念内容）和效标测量（criterion measurement，即效标的具体度量方法）。

2. 效标测量要求能真正反映效度，即它们的相关要高。另外，效标测量还必须具有较高的信度。

3. 效标测量的一个重要特性是客观性。首先要避免偏见的影响，尤其当效标测量是等级评定时，可能会受评定者印象或成见的影响。其次，应防止效标污染（效标污染指由于评定者知道测验分数而影响个人的效标成绩的情形）。

#### 4. 常见效标

由于每个测验的用途不同，因而会有不同的观念效标。常用的效标有以下几种：

##### （1）学术成就

这种指标常常作为智力测验的效标，也可以作为某些多重能力倾向测验和人格测验的效标。

##### （2）特殊训练成绩

能力倾向测验常用的效标是被试在将来某种特殊训练中所取得的成绩。

##### （3）实际工作表现

在许多情况下比较令人满意的效标测量是实际工作表现。这种效标可用于起选拔作用的一般智力测验、人格测验及能力倾向测验等。

##### （4）团体对照

采用团体对比法确定测验有效性的方法是：用两个在效标表现上有差别的团体，比较它们在预测源分数上的差别。

##### （5）等级评定

有时也可以采用其他人对应试者的效标行为的评定作为效标测量。这种评定往往由应试者的老师、同学、上级、同事等观察者做出。

#### (6) 先前有效的测验

一个新测验和先前有效测验的相关也经常作为效度的证据。

#### (三) 预测效度和同时效度

效标效度主要考虑测验分数与效标间的关系。根据效标材料收集的时间不同，可以将效标效度进一步区分为预测效度和同时效度。

1. 预测效度 (predictive validity) 的效标资料往往是测量结束后隔一段时间才获得，它反映的是由测验分数对任一段时间间隔后被试行为表现的预测程度。预测效度适用于那些对人员进行选拔、分类和安置的人事测验，这些测验需要对应试者将来的工作绩效进行可靠的预测。

预测效度的评估通常用追踪法来进行，通过长期观察，积累材料，以衡量测验结果对将来表现的预测能力。例如，可以对那些被测量的应试者进行长期观察，获得他们隔一段时间后的工作绩效的资料，看测验分数是否正确预测了他们的工作绩效的高低。

2. 同时效度 (concurrent validity) 的效标材料可以和测验分数差不多同时搜集。有时，同时效度可以替代预测效度。因为当测验施测于已存在有效效标材料的团体中时，就不必经过一段时间后再作比较。例如，大学生的测验成绩可以与其在学校的功课成绩相比较，选拔测验的得分也可以与应试者在现在工作中的绩效作比较。因为同时效度的评估不需要长期追踪，所以应用比较普遍。

同时效度和预测效度的差异的根源不是收集效标的时间，而是在于测验目的的不同。前者多用于诊断现在的状态，后者多预测未来的结果。这种差异可以用两种不同的询问方式说明：

**(1) “某人成功了吗?” “某人患病了吗?”**

**(2) “某人会成功吗?” “某人会患病吗?”**

第一类问题属于具有同时效度性质的测验要求回答的问题，而第二类问题则属于具有预测效度性质的

测验要求回答的问题，针对的是将来会发生的情形。

#### （四）确定效标关联效度的方法

1. 相关法：求测验分数与效标测量间的相关，称为效度系数。通常以斯皮尔曼或皮尔逊相关来表示。
2. 区分法：预测分数是否可以区分由效标测量所定义的团体。
3. 命中率。

（1）总命中率：正确决定数目对总数目的比率；

（2）正命中率：所选择的人成功的比例，选择效率。

#### 4. 功利率

$$U = B(N_s) - C(N_f) - S$$

$U$  表示功利率； $B$  表示成功录用带来的收益； $C$  代表错误录用带来的损失， $S$  表示程序的费用。

功利率的启发：测验如果简单易操作，省时省钱，不经过特殊训练的人可以掌握，适合团体施测，即使效度低一些，也可以采用。

#### 5. 合成法

- （1）工作分析，确定一项工作的基本元素以及相对权重；
- （2）确定每一个测验与相应元素的相关，估算测验与元素的相关；
- （3）依据元素的权重，将相关系数合成，得到合成效度。

## 第十章 测量的效度

### 第二节 效度的分类

#### 三、构想效度

##### （一）构想效度的概念

测验的构想效度 (construct validity) 又叫构想效度，是指测验能够测量到理论上的构想或特质的程度。



所谓构想通常指一些抽象的、假设性的概念或特质，如智力、创造力、言语流畅性、焦虑等。这些构想往往无法直接观察，但是每个构想都有其心理上的理论基础和客观现实性，都可以通过各种可观察的材料加以确定。例如，言语流畅性可以通过语速、语句间的逻辑性、口误的数量等可观察的指标进行确定。构想效度关注的问题是：测验是否能正确反映理论构想的特性。比如说，一项言语流畅性测验所测量的是不是真正的言语流畅性，是否对言语流畅性的理论概念中所包含的所有特点（如语速、语句间的逻辑性、口误的数量等）都进行了测量。

## （二）确定构想效度的步骤

一般而言，要确定一个测验的构想效度，包括三个基本步骤：

1. 建立理论框架，以解释被试在测验上的表现；
2. 依据理论框架，推演出各种有关测验成绩的假设；
3. 以逻辑和实证的方法来验证假设，根据这些累积材料决定这种理论是否能恰当地解释现有材料，

如果不能作出恰当解释，则应该修正上述假设，直到能作出恰当的解释为止。

构想效度的确定过程可以用一个简单的例子来说明。例如，某人对研究创造力这个构想感兴趣，他假设或认为那些具有创造力的个体与那些不具备创造力的个体有某些不同，因此可以建立一个理论（或一理论体系）来说明那些有创造力的个体（或具有创造力构想的个体）的行为与其他人的不同，从而使人们能够通过观察个体的行为和根据某种理论分类来辨别具有创造力的个体。如果希望编制一个测验来测量创造力，那么，这个创造力测验必须具有构想效度，即测验分数与根据创造力的心理学理论观察被试行为所作出的判断有相关。如果这种关系不成立，则该创造力测验缺乏构想效度的支持。如果关系不太大，可能有多种原因，例如，测验可能没有真正测量创造力，或者说关于创造力的理论是错误的。如果测验分数与根据理论作出的判断相关很高，表明测验具有构想效度。

由此可见，确定构想效度没有简单的逻辑分析或统计分析，而是从各种来源中逐渐累积资料以确定测验的构想效度。

## （三）常见的确定构想效度的方法

## 1. 测验内的方法

(1) 确定测验的内容效度：确定了测验的内容效度，有时候也就确定了构想效度。

(2) 考察测验的同质性和内部一致性。

A. 因素分析 (factor analysis)。因素分析是分析行为资料内部关系、结构特性的一种统计技术，比较适合于对构想效度的研究。通过因素分析可以找出测验中所包含的特质。卡特尔就是通过因素分析的方法从一系列形容词中归纳出 16 种共同因素，最终按照这 16 种特质构造成 16 因素人格测验的。

采用因素分析材料的构想效度分析过程可以这样进行：首先，对测验的所有项目进行因素分析统计，抽取出能够解释测验结果的大部分变异的一些共同因子，这些共同因子可以用来对测验的组成进行描述，然后比较由这些共同因子描述的测验组成是否与测量目标（某种特质或能力）的理论构想一致。

B. 考察总分数较高和总分数较低两类人在各个题目上通过率的大小。考察总分数较高和较低两类人在各题目上通过率的大小，比较每个题目上总分最高的 27% 的被试与总分最低的 27% 的被试，如果前者在该题上的通过率显著大于后者（一般要求在 30%~40% 以上），则认为题目是有效的，否则应淘汰或修改。

C. 计算题目与总分相关。

D. 计算分测验与总分相关。

E. 其他计算同质性信度和内部一致性信度的方法。

(3) 分析被试对题目作反应的过程。

## 2. 测验间的方法

(1) 相容效度：测量相同特质或构想的测验，彼此之间应该有高相关。因此一个新测验与相似的旧测验之间的相关，可以作为衡量新测验所大致测量的相同行为的程度的标准。这种相关系数有时又称作相容效度。与效标效度不同，这类相关较高，但不很高。如果一个新测验与现有的有效测验的相关很高，而且不是更简便或易于实施，那么就没有必要编这个测验。与其他测验的相关还有另一种用途，就是表示新测验是否受到某些无关因素的影响。一般来说，测量不同构想或特质的测验之间，相关要很低，因此，一个能力倾向测验不应该和人格测验有高相关，学术能力测验不应该与管理技能测验有太高相关。

(2) 区分效度与会聚效度：1960 年坎贝尔 (D. T. Campbell) 指出，要确定一个测验的构想效度，则该测验不仅应与测量相同特质或构想等理论上有关的变量有高的相关，也应与测量不同特质或构想等理论上不同的变量有低的相关，前者称为会聚效度 (convergent validity)，后者称为区分效度 (discriminate validity)。这就是坎贝尔和费斯克 (D. W. Fiske) 1959 年提出的构想效度的一种考验方法。例如，一个数学推理能力测验与数学课成绩的相关就是会聚效度；而该测验与阅读理解能力测验的相关显著低，就是区分效度，因为在测验设计时，已将阅读能力作为数学推理能力的无关因素来考虑。区分效度特别适合于人格测验，因为人格测验比较容易受到各种无关变量的影响。

(3) 多特质——多方法矩阵。

### 3. 研究测验的实证效度

(1) 如果一个测验有实证效度，那么该测验所预测的效标的性质与种类就可以作为测验所要测量的构想的指标，研究这些资料就可以提供与构想效度有关的信息。

(2) 测验对效标的预测能力，可以提供构想效度的证据。

### 4. 考察实验变量对测验分数的影响

通过实验操作某些条件，观察其对测验分数的影响以及测验分数与某些效标之间的关系，可以获得构想效度的信息。如，考试焦虑——当考试结果对个人有重要意义时的一种害怕失败的紧张情绪。方法 1：焦虑分数在大考和一般考试前的比较；方法 2：将考试分数作为效标；方法 3：将生理指标作为效标。

### 5. 收集某些变异的证据

根据构想效度分析变异的变化趋势，以确定某些构念的构想效度。

变异的证据是必要条件，而非充分条件。例如，某些智力和技能。

## 第十章 测量的效度

### 第三节 效度的影响因素

#### 一、测验的因素

凡是能造成测验结果误差的因素，都会影响测验的效度。一个测验的效度高低，很大程度上取决于该测验受无关因素影响的程度。受无关因素影响越小，则效度越高。由测验本身带来的影响因素有测验题目的质量、测验实施中的干扰、测验的长度、被试的因素等。

### （一）测验题目的质量

题目的指导语不明确、试题的表达不清晰、试题太难或太容易、题目中出现额外的线索、诱答设计不合理、题目过少、试题的安排和组织不恰当、试题不符合测验目的等因素，都会影响测验的效度，使效度降低。

### （二）实施测验时的干扰因素

测验的环境太差、被试不遵从指导语、计分错误，都会使测验的效度降低。对于效标效度，效标获取的时间与测验时间相隔越长，测验结果与效标的关系受无关因素的影响就越大，所求得的效度必定越低。很显然，这些因素使得测量分数反映的不单是测量内容，而且是无关因素的干扰。细心的读者会注意到，这些因素同样也会影响信度，因为它们会使测量结果波动不定。

### （三）被试的影响因素

被试的反应定势、测验动机、情绪和身心状态都会对测验的结果造成影响，所以这些也将会影响测验效度。

### （四）测验的长度

一般来说，增加测验的长度通常可以提高测验的信度 $r$ ，而效度系数能否达到最大值也受信度的影响，因此，增加测验的题目往往也能提高测验的效度。不过，效度增加的前提是这些增加的题目必须与测量的目标相关。

测验长度对效度的影响可以用下面的公式来表示：

$$\eta_{xx}^2 = \frac{nr_{xy}}{\sqrt{n(1-r_{xx} + nr_{xx})}}$$

其中， $\eta_{xx}^2$ 为测验增长到原来的 $n$ 倍后，测验 $(x)$ 与效标 $(y)$ 的相关（即增长后的效度系数）， $n$ 为测

验增长的倍数； $r_{xy}$  为原测验的效度系数； $r_{xx}$  为原测验的信度系数，根据公式可以推知，要达到满意的效度水平，测验题目需要增加几倍。

## 二、样本团体的性质

对效度的计算往往是通过样本团体的分数进行各种分析而得到的，所以样本团体的性质也会对测验的效度产生影响。这种影响体现在三个方面：

（一）同一测验对不同的团体所测量的功能可能是不同的。

例如，同一项算术测验，对于能力较差的被试可能测量的是数学推理能力，但对于能力强的被试，可能测量的只不过是以前所学内容的回忆能力和计算能力。在评价效度时，我们要力求使样本团体的性质与所要测量的团体的性质尽量相似，这样所求得测验效度才会较高。

（二）对于同一个测验，效度团体的性质不同，效度也会有较大的差别。

样本团体的性质包括年龄、性别、教育水平、智力水平、动机水平、职业等有关特性。同一测验对不同性质的团体可能有不同的预测能力，因此我们将这些对测验的效度产生影响的因素称为干涉变量。例如，对出租汽车司机实施能力倾向测验后发现，测验成绩与司机的工作成绩的相关仅达 0.20，这是相当低的预测效度。但是，当把对开车有兴趣的司机抽出来单独计算效度时，却发现效度是 0.60。这说明该测验虽然对所有的司机的工作成绩没有预测能力，但对于“有工作兴趣的司机”这一亚团体却有较好的绩效预测力。在这个例子中，兴趣就是一个干涉变量。

（三）样本团体的异质性对效度也会有影响。

用相关系数表示的效度系数会受到取样团体的分数分布的影响。如果其他条件相等，那么样本越同质，效度越低。反之，效度越高。

## 三、效标的性质

在采用效标关联效度时，效标的性质如何，会影响对测验效度的评价。一般来说，如果其他条件相同，所测量的行为或心理特质与效标行为或特质越相似，效度系数就越高。比如，假设我们要建立一项管理能力测验，现在要为此测验选定效标。也许有人会考虑采用下列效标来源：学历、升入现职位的速度（年

限)、近三年管理工作的业绩、同行的评价、上司的评价等。显然,这些效标与测验内容的同质性并不相同,将测验分数与它们求相关时,得到的效度值大小也就不同。

另外,效标与测验分数之间的关系是否线性也是一个很重要的影响因素。因为,皮尔逊积差相关的前提假设是,两个变量的关系是线性的均匀分布。如果测验分数与效标之间的关系是非线性的,采用皮尔逊积差相关将会低估相关的大小,造成效度的低估。

# 全国迷你型MBA职业经理双证班

**认证系列:** 职业经理资格认证、人力资源总监、营销经理、品质经理、生产经理、物流经理、项目经理、企业培训师、酒店经理、市场总监、财务总监、营销策划师等认证。

**颁发双证:** 通用高级经理资格证书+MBA 高等教育研修结业证书(含2年全套学籍档案)

**证书说明:** 证书全国通用、国际互认、电子注册,是提干、求职、晋级、移民的有效依据

学习期限: 3个月(允许工作经验丰富学员提前毕业) 收费标准: 全部学费 **1280** 元

咨询电话: 13684609885 0451- 88342620 88723232 电子邮箱: [xchy007@163.com](mailto:xchy007@163.com)

学校网站: [www.mhjy.net](http://www.mhjy.net) 颁证单位: 中国经济管理大学 承办单位: 美华管理人才学校

**全国招生 函授教育 颁发双证 权威有效**



**职业经理 MBA 整套实战教程**

千本好书 **免费** 下载 学校网址 [www.mhjy.net](http://www.mhjy.net)

## 第十一章 项目分析

### 第一节 项目的难度分析与诱答分析

#### 一、项目的难度分析

### （一）难度分析的概念

难度，通俗地说就是指题目的难易程度。对能力测验来说，难度就是指测验题目是难还是容易，这种测验往往有一个反映难度水平的指标。对于非能力测验（如人格测验），难度是指测验题目的“通俗性”，即测验题目是否容易被别人看懂和回答，往往以取自相同总体的样本中，能在答案范围内回答该题的人数指标来表示。难度分析（difficulty analysis）就是对题目的难度进行估计以确定适宜的难度。由于非能力测验的难度分析相对简单，而且往往可以通过测验编制者的经验来把握，所以下面不对非能力测验的难度分析作具体介绍，而主要介绍能力测验的难度分析，并与非能力测验的难度分析相对参照。

### （二）难度分析的方法

难度的指标通常以通过率表示，即以答对或通过该题的人数百分比来表示：

$$P = \frac{R}{N} \times 100\%$$

其中， $P$  为试题难度； $N$  为全体被试人数； $R$  为答对或通过该题的人数。

以通过率表示难度时，通过人数越多（即  $P$  值越大），题目越容易，难度越低； $P$  值越小，题目越难。

#### 1. 难度的计算

若试题为二分法反应一计分项目（即答对给 1 分，答错记 0 分）， $P$  值可以直接采用上面的公式计算。当被试人数较多时，则可以根据测验总成绩将被试分成三组：分数最高的 27% 被试为高分组（NH），分数最低的 27% 的被试为低分组（NL），中间 46% 的被试为中间组。分别计算高分组和低分组的通过率，再求试题的难度。

$$P = \frac{P_H + P_L}{2}$$

或

$$P = \frac{1}{2} \left( \frac{R_H}{N_H} + \frac{R_L}{N_L} \right)$$

公式中， $P$  代表难度； $P_H$  和  $P_L$  分别代表高分组和低分组通过率； $R_H$  和  $R_L$  分别代表高分组和低分组通过该题的人数； $N_H$  和  $N_L$  分别代表高分组和低分组的人数。

**例如：在 100 名被试中，选为高分组和低分组的被试各有 27 人，其中高分组有 20 人答对第一题，低分组有 10 人答对第一题，则第一题的难度为：**

$$P = \frac{1}{2} \left( \frac{20}{27} + \frac{10}{27} \right) = \frac{0.74 + 0.37}{2} = 0.56$$

## 2. 选择题的难度修正

在采用多项选择题的测验中，由于猜测的因素，被试的得分可能会比被试的真实得分要高。并且，选项的数目越小，机遇的作用越大，被试的能力被夸大的可能性就越大，就越不能真正反映测验的难度。为了平衡选项数目对难度的影响，吉尔福德提出了一个难度的校正公式：

$$CP = \frac{KP - 1}{K - 1}$$

公式中， $CP$  为校正后的通过率； $P$  为实际得到的通过率； $K$  为选项的数目。

## 3. 非二分法记分的项目

$$P = \frac{\bar{X}}{X_{\max}} \times 100\%$$

其中  $\bar{X}$  代表在该题上的平均分， $X_{\max}$  代表该题的满分。

### （三）等距量表的难度指数

以通过率作为难度指标，实际上是以顺序量表来表示难度，即它仅仅能指出题目难度的顺序或相对难度的高低。例如有 3 个试题，第 1、2、3 题通过的人数分别为全体被试的 60%、40%、20%，我们可以判定三个项目中，第 1 题最容易，第 3 题最难。但是虽然它们的相差数值相同（20%），我们却无法确定题目 1 与 2 之间的难度差别是否等于题目 2 和 3 之间的难度差别。在需要比较题目差异的情况下，就需要计算等距量表的难度指标。



如果被试的得分呈正态分布，则可以根据正态曲线表，将试题的难度转换成具有相等单位的等距量表，即分数。例如，我们知道，在正态分布情形下，平均数之上或之下一个标准差的距离约占全体人数的 34%。因此，如果在一个测验中，某题 A 通过率为 84% ( $p=0.84$ )，从正态分布表中可以知道该题的难度为  $-1\sigma$ ；如果某题 B 的通过人数只有 16%，则这个题目的难度为  $+1\sigma$ ；如果某题 C 刚好有 50% 的被试通过，那该题的难度为 0。显然，较难的题目难度为正值，简单的题目为负值。根据正态曲线表，我们可以查出任何通过率值相对的  $\sigma$  值。

以分数表示难度，也有不便之处，主要是有小数点和正负符号，使结果不易处理。因此通常需要转换成另一种单位的等距量表，其中较为常用的是美国教育测验服务机构采用的难度指标：

$$\Delta = 13 + 4Z$$

公式中， $\Delta$ （希腊字母 delta）为正态化等距难度值；13 为平均数；4 为标准差的单位， $Z$  为正态化  $Z$  分数值。

**例如，上例中，题 A 的通过率为 0.84，则  $Z$  值为  $-1\sigma$ ，则  $\Delta = 13 + 4 \times (-1) = 9$ ，题 B 的难度值为 0.16， $Z$  为  $+1\sigma$ ， $\Delta = 13 + 4 \times 1 = 17$ ，题 C  $\Delta = 13$ 。如果一个题目几乎所有被试都通过 (99.8%)，则其  $Z$  值大  $-3\sigma$ ， $\Delta = 1$ ；如果某题答对的人数极少 (0.13%)，则  $Z$  值为  $+3\sigma$ ， $\Delta = 25$ 。因此， $\Delta$  量表是以 25 为上限，1 为下限的等距量表， $\Delta$  值越大，则难度越高， $\Delta$  值越小，难度越低。**

对大多数测验而言，只要算出  $P$  值即可，但如果要作精确的统计分析，则需要计算出具有等距性质的  $\Delta$  值。

#### （四）测验难度的确定

##### 1. 题目难度水平的确定

题目难度水平究竟多高合适，依赖于测验的目的、性质和题目形式。

##### （1）从测验的目的考虑

许多测验都希望能准确测量个体间的差异。如果在某题上，被试全答对或全答错，则该题无法提供个别差异的信息，即不能把不同人区别开来。 $P$ 值越接近于 0 或接近于 1，越无法区分被试能力的差异。相反， $P$ 值越接近于 0.50，区别力越高。道理很简单，如果 100 人中有 50 人通过，则共会有 2500 个配对比较信息 ( $50 \times 50 = 2500$ )；如果 30 人通过，则有 2100 个比较信息 ( $30 \times 70 = 2100$ )；而 90 人通过时只有 900 个比较信息 ( $90 \times 10 = 900$ )。

为了使测验具有更大的区别力，应选择难度在 0.50 左右的试题比较合适。但是在实际工作中并非如此简单。如果难度都是 0.50，则试题间的相关将有偏高趋势。举一个极端例子，假定某测验各试题间相关均为 1.00，题目难度均为 0.50，那么有可能使 50% 的被试答对所有题目得满分，另外 50% 的被试无法通过任何试题，而全部得 0 分。这样仍起不到最大限度区分所有被试的效果。所以在选择题目时，最好使试题的平均难度接近 0.50，而各题难度在  $0.50 \pm 0.20$  之间。

## (2) 从测验的作用考虑

对于选拔用的人事测验，应该尽量使难度值接近录取率。比如，测验若是要辨别或选择少量最优秀的应试者，测验就应该有比较高的难度， $P$ 值应该较小。如果录取率为 5%，那么题目难度最好也确定为 5%，使得恰好能使 5% 的优秀人选通过；假如测验是要筛选出少数较差的被试，则题目  $P$ 值应该高，使得只有少数应试者不能通过。

## (3) 从题目的形式考虑

对于选择题而言， $P$ 值一般应大于概率水平。 $P$ 值等于概率，说明应试者可纯粹凭猜测回答。 $P$ 值比概率还小，说明题目有严重质量问题，即被试即使靠猜，或者说不需要猜，都能答出题目。吉尔福德公式是选择题难度修正公式，我们知道真正的难度为 0.50 时最合适 ( $cp = 1/2$ )，那么得到下面的公式：

$$P = \frac{K+1}{2K}$$

它表示对于具有某种选项数 ( $K$ ) 的项目，其较为理想的应得难度值。对于是非题而言，其难度值应该为 0.75 最为合适；而对于四选一题目，其难度值约为 0.63 时最为合适。

## 2. 测验难度水平的确定

整个测验的难度取决于组成测验的各个试题的难度。整个测验难度水平的确定，需要根据测验分数的分布作出。由于众人的心理特性基本上是呈正态分布的，因此如果被试样本具有代表性，则测验结果也应符合正态分布的模型。

当然，也不是所有测验都要求测验分数呈正态分布。有些效标参照测验，出现偏态分布是允许的。

## 二、项目的诱答分析

### （一）诱答分析的概念

诱答就是指题目中除正确选项外的其他错误选项。测验题目的诱答设计得如何，直接关系到测验题目的好坏、测验的难度、鉴别力以及测验的信度和效度。诱答分析（distracter analysis）是指系统考察测试题目，以诊断题目是否符合一系列测量学要求的过程，主要目的在于避免猜测的影响，使题目真正反映应试者的真实情况。

### （二）好诱答的标准

1. 诱答分析通常针对能力测验。对于能力测验而言，其测验目的无非是对能力强和能力弱的被试进行区别。这就要求测验题目对于那些能力强、掌握正确知识的被试来说有更大把握选择到正确答案，而那些能力差、没有掌握充分知识的被试往往选择错误答案。另外，对于一个完美的诱答设计，应该使不知道正确答案的被试对题目各个选项随机选择，每个选项被不知道答案的被试选择的概率相同。例如，对于一个四择一的选择题，每个项目被不知道答案的被试选择的概率都应该是 25%。

2. 对于以选择题为形式的人格测验来说，并没有“诱答”之说。但是对各个选项的分析也是必要的。在人格测验，尤其是在是自陈量表的人格测验中，选项的设计要力求避免社会赞许倾向。如果测验中出现社会赞许倾向高的选项，具有不同人格特点的被试都会倾向于选择这个选项，测验就无法达到测量个体差异、反映个体真实情况的目的。因此，在人格测验中，选项设计不好同样会导致测验的鉴别力和效度的降低。

3. 以是非题为形式的测验，可以将是非题看成只有两个选项的选择题，同样可以进行诱答分析。

### （三）诱答分析的方法

1. 进行诱答分析的最一般的方法，就是分别计算每个选项被作为正确答案选择的次数。
2. 理想的诱答设计方法是使每个诱答选项的百分比相等。

每个诱答被选择的理想百分比=答错的百分比/诱答数目。

### 3. 设计诱答时的注意事项

(1) 不要出现过于容易被排除的诱答，这会使不知道正确答案的被试猜测正确的概率提高。

(2) 不要出现过于似真的选项，即欺骗性诱答。如果被试掌握了题目所要考察的全部知识仍很难对正确答案与某一诱答进行区分，那么该诱答就是欺骗性诱答（deceiving distracter）。欺骗性诱答的出现使题目实际考察了目标之外的能力，使测验的效度降低。

## 第十一章 项目分析

### 第二节 项目的鉴别度分析

#### 一、项目鉴别度的概念

项目鉴别度（item discrimination）是指测验项目对于所测查的心理特性的鉴别能力和区分程度。鉴别度分析可以从项目效度和内部一致性两方面着手。

#### 二、项目效度分析

项目效度分析（item validity analysis）主要以效标为依据，考察被试在每个试题上的反应与其在效标上表现的相关程度，即每个试题所测查的行为是否反映了被试在效标上的表现。通常，鉴别度分析的指标有相关系数和鉴别度指数两种。

##### （一）相关系数

这里的相关系数指项目通过率与效标成绩的相关系数。相关越高，表明项目越具有区分的功能。这一相关系数的计算方法有很多，根据不同的情况采用不同的计算方法。事实上，用来表示项目效度系数的方法至少有 50 种，虽然方法不同，但所得结果基本上是一致的。至于采用何种方法计算最佳，以方便为标准。题目效度系数越高，对预测效标越有用。效度系数高于 0.20 的题目一般可以保留；当效度系数接近于

零或负数，则应该作重大修改或删除。下面列举了通常的几种情况和常用公式：

1. 通常情况下，项目的分数以二分法计分，如答对计 1 分，答错计 0 分，而效标成绩可能是连续的分数。求此二者之间的相关，应采用点二列相关系数：

$$r_{pb} = \frac{(\bar{X}_p - \bar{X}_q)}{S_t} \sqrt{pq}$$

公式中， $\bar{X}_p$  代表通过该题被试的平均效标成绩； $\bar{X}_q$  代表未通过该题的被试的平均效标成绩； $p$  代表通过该题的人数百分比； $q$  为未通过该题的人数百分比； $S_t$  代表全体被试的效标成绩的标准差。

2. 如果假定被试在试题上的反应也是正态分布，只是由于人为方法分成通过或不通过两种情形，则可以采用二列相关公式：

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{S_t} \frac{pq}{y}$$

公式中， $y$  是正态分布下答对百分比( $p$ )所在位置的曲线的纵轴值。

3. 在有些情况下，对没有通过某题被试的效标成绩不感兴趣时，则可用总人数的结果进行计算，上面的两个公式可以转换成：

$$r_{pb} = \frac{(\bar{X}_p - \bar{X}_t)}{S_t} \sqrt{\frac{p}{q}}$$

$$r_b = \frac{(\bar{X}_p - \bar{X}_t)}{S_t} \frac{p}{q}$$

公式中， $\bar{X}_t$  为所有被试的平均效标成绩。

4. 当题目与效标均是人为二分的项目时，可以得到四个分数类别：即题目和效标均通过的 A 类；题目通过、效标不通过的 B 类；题目不通过、效标通过的 C 类，和题目、效标均不通过的 D 类。计算这种相关常采用皮尔逊余弦  $\pi$  公式，所得相关为四分相关系数( $r_t$ )：

$$r_t = \cos\left(\frac{180^\circ}{1 + \sqrt{AD + BC}}\right)$$

公式中，A、B、C、D 分别代表每类包含的次数。查三角函数表，可得余弦相关系数。

5. 当两个变量都是二分变量时，亦可计算  $\theta$  相关系数：

$$\theta = \frac{AD - BC}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

## （二）鉴别度指数

鉴别度指数 (discrimination index)，它是比较效标得分高和得分低的两组被试在项目通过率上的差值：

$$D = P_H - P_L$$

当效标成绩是连续变量时，可从分数分布的两端各选择 27% 的被试组成高、低分组，再分别计算两组每道题目上各自的通过率，二者之差就是各题目的鉴别度指数  $D$ 。 $D$  值是鉴别题目测量效标有效性的指标， $D$  值越高，题目越有效。当  $D=1.00$  时，高分组全部通过，低分组全部失败。当  $D=0$  时，则高分组和低分组通过的人数相同。在一般情况下， $D$  值很难等于 1.00，而且也没有必要甚至不能追求等于 1.00，因为这种情况往往只是题目分数呈现绝对的双峰分布时才会出现。这时，各题目的同质性也会过高。所以，一般情况下，只要所得  $D$  值大于 0.30 时，题目就可以接受了。

鉴别度指数  $D$  与前面提到的项目分析指标  $r_b$ 、 $r_{pb}$ 、 $r_t$ 、 $\theta$ ，虽然单位、计算公式和数值不同，但其结果和结论却有很高的正相关，因此  $D$  可作为上述系数的简便替代方法。目前采用的项目分析法，常用鉴别度指数为依据。

另外，鉴别度  $D$  值和难度值  $P$  也不是相互独立的指标。当  $P$  值为 1.00 或 0 时，则高分组和低分组通过的百分比完全相同，因此  $D$  值必然为 0；而  $P$  值为 0.50 时，则可能是高分组全部通过，低分组全部失败， $D$  值可能达到 1.0。因此难度  $P$  值处于中等水平时， $D$  值可以达到最大。

## （三）高、低分组的划分

在题目难度和鉴别度分析中，多次提到要划分高分组和低分组。一般情况下，这一分组是根据效标成绩或总成绩将被试排队，从高分端取 27% 的被试组成高分组，从低分端取 27% 的被试作为低分组，其余的 46% 被试可以不作分析。当效标分数分布是正态分布时，这种划分标准很合适。但是，当分数分布较正态分布平坦时，高、低分组各占比率应该有所调整，一般来说应该高于 27%，大约 33% 较为合适。一般高、低分组各占比率在 25%~33% 之间都可以。如果比率太小，则所选出来的两组过于极端，更容易有明显区别，故可能是人为夸大了题目的区分力。当样本团体过小时，划分的比例可以适当提高，甚至用 50% 的标准，即把上下两半被试作为高分组和低分组都是可以的。

### 三、内部一致性分析

（一）有时在进行项目分析时，难以找到合适的效标，这时可以采用测验的总分代替效标，考察每个试题和总分的一致性，即做项目—总分分析（item-total correlation analysis）。这种分析反映的是测验各项目所测查内容的一致性，而不能反映题目对效标的有效性。

（二）内部一致性的计算方法与项目效度相同，只不过用总分替代外在效标成绩。第一种方法是求试题得分与总分的相关程度，仍然可以采用前述的点列相关和二列相关系数，以表示内部一致性高低。第二种方法是比较总分高低两组在每道题上通过率的差异，求得鉴别度指数高低。 $D$  值越大，即表示该题能对总分高低不同的被试作出区别。

（三）把项目和总分相关作为项目鉴别度的指标需要满足一个条件，那就是要求每个项目的方差都相同，或者是测验项目较多，各题所贡献的方差比例都较小。因为在计算项目与总分的相关时，有一个局部和总体的相关问题。在总分方差中实际包含有各个试题的方差，这本身就有可能造成某种相关。如果各个题目对总体方差的贡献差异显著，会造成某些题目与总分高度相关，一些题目相关却很低的情况，而这种相关差异是由题目方差不同造成的。

如果不具备上面的条件，最好是计算每题与所有其他项目合成分数之间的相关。这个合成分数可以采用下面的公式进行估计：

$$r_{pq} = \frac{r_{tp}S_t - S_p}{\sqrt{S_t^2 + S_p^2 - 2r_{tp}S_t \cdot S_p}}$$

其中， $r_{pq}$ 为某一项目与所有其他项目合成分的相关； $S_t$ 为总分的标准差； $r_{tp}$ 为某一项目与测验总分的相关； $S_p$ 为项目的标准差。

#### 四、项目效度与内部一致性的比较

项目效度和内部一致性分析虽然步骤相同，但意义却有差别。在筛选试题时，一定要了解这两种分析方法的意义和性质，这样才能正确地筛选题目，使测验达到预期效果。

项目效度代表题目与外在效标的关系，内部一致性则更多地代表题目与总分的关系。有较高的内部一致性的题目，并不一定与外在效标有很高相关。同样，部分具有很高项目效度的题目，与测验总分的相关并不理想。例如，某项能力倾向测验由 80 道逻辑推理题和 20 道数量分析题组成。如果以内部一致性作指标，则由于逻辑题在总分中所占比重较大，结果显然是逻辑题与总分的相关要高。但是如果这项能力倾向测验既想考察应试者的逻辑推理能力，又需要考察其数量分析能力，那么只选择内部一致性高的题目，就会造成所选题目全是逻辑题，改变了测验的初衷。因此，项目效度和内部一致性有时候是有矛盾的，我们在进行分析时一定要全面地考虑问题。

#### 五、鉴别度的标准

鉴别度的值具有相对性，采用不同的计算方法或不同的分组标准所得到的鉴别度的值是不同的。另外，鉴别度还与受测团体的同质程度有关；所测团体同质性越强，鉴别度就越小。但若施测于较为异质的团体，则鉴别度就有可能提高。假设有一个心理学知识测验，施测于心理系学生（团体较同质），则鉴别力可能较小，但若施测于一般大学生，鉴别度可能就较大。

由于鉴别度的相对性，很难确定一个筛选项目的绝对标准。因为除了考虑采用哪种区分度指标，以及样本大小和特性外，还要考虑测验的目的、性质和功能。就成就测验而言，一般要求项目和总分的相关达到 0.20 以上，而鉴别度指数希望达到 0.15-0.20 以上。一般心理测验， $D$  值在 0.30 以上就可以接受了。



### 第三节 其他类型的项目分析

#### 一、速度测验的项目分析

##### （一）速度测验的概念

有些测验考察的是被试解决问题的速度。这样的测验叫速度测验（speed-based test）。速度测验往往有很多题目，一般很少有被试能够在限定时间内完成所有的题目。大多数能力测验都限定了完成时间，不是所有被试都能在限定时间内完成所有项目。虽然这些测验不是速度测验，但速度也是影响成绩的一个重要因素。

##### （二）速度测验的项目分析方法

1. 速度测验的项目分析不能直接采用上面提到的各种方法。因为往往只有一部分被试回答了测验较靠后的题目，这些题目一般都不太难，只是被试的答题速度太慢，没有时间回答。如果按通常的方法，用通过率作为项目分析的指标，势必会造成对这些题目的难度过高估计。同样，由于能够做后面题目的被试的能力往往比其他被试要高，以鉴别度指数（高分组通过率减低分组通过率）来分析往往会造成对后面题目的鉴别度高估的现象。看来，通常的项目分析的方法不能用于速度测验。

2. 对于速度测验的项目分析，有人提出两种方法来分析。

（1）对于那些不能被所有被试完成的题目只分析完成该题目的被试的成绩。例如，一项速度测验共有 100 道题，试测的 100 名被试中只有 80 名完成了该测验的第 78 题，其中回答正确的被试为 72 名。那么按上面的方法计算，该测验的第 78 题的难度指数为： $72/80 = 0.9$ 。但这种方法也有不足之处。首先，如果假定能够完成第 78 题的被试的能力通常较高，那么这部分被试在该题上的通过率必定较高，这就造成对该题难度的低估。其次，采用这种方法，越处于测验后部的题目其数据样本越小，统计出的难度指数就越不稳定。另外，由于各个题目所采用的样本不同，无法进行前后各题目之间的难度比较。

（2）延长施测时间，使所有被试都能完成所有题目，然后对所有被试的成绩进行分析。这种方法适用于那些不以速度作为测试目的，但速度影响测试结果的难度测验。但如果测量目标中包含速度因素，这种方法就不适合了。总之，对于速度测验的项目分析尚没有完善的解决办法。

解题速度快慢也是能力大小的一种表现，所以有些能力测验中往往包含速度因素，不是所有被试都能完成全部题目。对于这种能力测验的项目分析，要注意以下两点：第一，将试题按估计的难度由易到难的顺序排列，这样即使延长让能力差的被试完成后面题目，他们也往往会回答错误；第二，在施测时，不妨适当延长答题时间，尽量让所有被试完成全部题目。

## 二、效标参照测验的项目分析

### （一）效标参照测验不适用于前面所说的项目分析方法

前面所谈到的项目分析方法通常只适用于常模参照的测验，而对效标参照测验并不适合，如成就测验。在考察培训效果时，往往会用到成就测验。例如，公司举办了一个培训班，对刚进入企业的新员工进行培训。培训结束后，人事经理想要了解培训班的效果如何时，往往会对经过培训的人员进行一次测试，看员工是否比培训前有所进步。这时，会遇到一个问题，这次测试的题目是否能够真正体现对培训效果的考察。如果测试的题目测试了不属于培训内容的能力，这项测验就不成功。这种情况下，即便该测验能够对能力高低的员工进行很好的区分，也不能采纳，因为这项测验内容没有很好地反映它的效标——培训任务，这就是效标参照测验与常模参照测验的区别。

对效标参照测验的项目分析不能采用上述方法。因为效标参照测验的难度完全取决于效标的难度。例如，如果培训内容很难，那么对培训效果进行考察的测验也必定较难；反之，如果培训内容本身很简单，那么测验也必定简单。另外，对于效标参照测验，鉴别度不能说明问题。例如，当培训内容十分简单时，测验的题目也必定简单，这种情况下可能所有的被试都能答对题目，即试题的鉴别度为零。但是，我们不能由此得出该题不好的结论，因为试题实际上的确考察了培训内容，能够帮助人事经理了解培训效果。

### （二）效标参照测验的项目分析方法

1. 有人用对前后两次施测结果进行比较的方法来确定测验题目的好坏。例如，在培训之前，对所有人员施测该测验。在培训结束之后再次施测。如果某一题目在培训后的平均得分高于培训前，则说明该题目反映了培训的内容和效果，是一个好的试题。这种方法应用较普遍，但也存在一些问题。首先，前后两次施测需要花费较多精力和时间；其次，存在重复测验引起练习效应的问题；另外，如果培训后的成绩等

于或低于培训前，不能说明该试题不好，有可能是教师的教学效果不佳。

2. 有些研究者用“教育敏感性”指标来进行分析。教育敏感性（education sensitivity）是指一种教育培训项目能否产生特定效果的可能性程度。选取两组被试（人数相同），一组进行培训，另一组不参加培训。培训结束后，对两组被试施测待分析的测验。计算每个题目的“教育敏感性”指数：

$$S = \frac{R_H - R_L}{N}$$

其中， $R_H$ 为受培训被试中答对该题的人数； $R_L$ 为未培训的被试中答对该题的人数； $N$ 为总人数。

用这种方法进行项目分析要注意两组被试的匹配的问题。所选取的两组被试必须在能力、知识背景等方面没有差别，要保证两组被试的各方面尽量相同。

## 第十二章 测量工具的建立

### 第一节 针对不同目的设计测量工具

#### 一、测量目标的分析

##### （一）测量方向确定

测量目标是指测量工具作什么用。测量目标不同，所设计的测量内容、测量工具的操作形式、采用题目的形式、题目的范围和难度都会有差异。

在确定测量内容之前，先要确定测量目的是显示个体的行为特点，还是用来预测其将来的行为表现。例如，成就测验、态度测验就是显示性测量（display measurement）工具，它反映被测者具有什么知识和特点，能完成什么任务。相对而言，许多观察法、调查法都是显示性测量工具。

预测性测量（forecast measurement）工具帮助我们预测一个人在不同情境下的行为。例如，各种能力倾向测验可以帮我们预测个体未来的工作绩效。部分人格测验可以预测个体未来的工作风格。而角色扮演、公文筐测验更加能够直接预测个体未来在同样的工作情境中的工作表现。

##### （二）目标分析

在实际过程中，测量工具的编制者还必须把测量目标转换成可操作的测量指标，这种转换过程我们称之为操作化或目标分析过程（purpose analysis）。

目标分析与测量目标是密切相关的。根据测量目的的不同，我们区别以下三种情况来具体分析：

1. 对于有选拔和预测功能的预测性测量工具，它的主要任务是要对所预测的行为活动做具体分析，我们称之为任务分析（task analysis）或工作分析。这种分析包括两个步骤：

（1）要确定使所预测的活动达到成功所需要的心理特质和行为。例如职业能力倾向测验的编制，若某项工作包括打字，那么测量工具的编制者可以假定手指的灵活性、手眼协调等能力是必需的。这种确定可以通过参阅前人的工作从理论上分析，也可以通过对在某项活动中已经录用或已经成功的从业人员的行为进行分析后得出。当测验编制者确定某项工作需要哪些能力、技能或特质之后，他就可以编制测量这些能力或特质的测量工具。

（2）建立衡量被试成功与否的标准，这个标准我们称之为效标（criterion）。例如，用以确定一名运动员是优秀运动员的标准，就是效标。效标可以作为鉴别测量工具的预测是否有效的重要指标。

2. 如果测量工具用于测量一种特殊的心理品质或特质，那么首先就必须给所要测量的心理和行为特质下定义，然后找出该特质往往通过什么行为表现出来。例如创造力的测量，有人将创造力定义为发散性思维的能力，即对规定的刺激产生大量的、变化的、独特的反应的能力。根据这个操作性定义，创造能力则应该从反应的流畅性、灵活性、独创性和详尽性这四个方面来测量。

3. 如果测量工具是描述性的显示测验，它的目标分析的主要任务则是确定所要显示的内容和技能，从中取样。成就测验就是一种典型的描述性显示测验，它的内容分析可以利用双向细目表来完成。双向细目表（two-way checklist）是一个由测量的内容材料维度和行为技能维度所构成的表格，它能帮助成就测量工具的编制者决定应该选择哪些方面的题目以及各类题目应占的比例。如果要检查新员工培训班的培训效果如何，就需要采用成就测验。这种成就测验的内容就可以利用双向细目表来确定。表 12—1 是为测量学生学习“测量工具的编制和实施”这一教学内容中的成就所编写的测量工具双向细目表的一个样例。

表 12—1

“测量工具的编制和实验”课程内容的测量工具细目表

行为 目标	内容（材料）				
	测量工具准备 （16%）	测量工具编表 （20%）	实验 （14%）	记分 （22%）	项目分析 （28%）
概念 知识	工作分析 效标事件 显示性测量工具 （3题）	四配型 混合性量表 反应定势 （5题）	关系效应 量标效应 （2题）	记分能 别合记分 机器记分 （3题）	效标 内每一事件 测量工具同质性 （3题）
具体 事实 知识	教育目标分类的 类别 （2题）	问答题和客观题 的优缺点和点 （4题）	测量工具造成的 影响因素 （2题）	问答题过分和客 观题记分的规则 （3题）	决定项目的效度 的方法 项目分析的目的 （3题）
理解	编制测量工具 计划的目的 （2题）		标准化	记分起加权的作用 （1题）	P和D关系理解 （1题）
应用	编制测量工具 时某一单元的效 内细目表 （1题）	测量理解、应用、 分析综合、评价的 多项题实例 （4题）	编制测量工具指 示语 （2题）	测验的修正 力加权 使用图表对重排 题记分 （4题）	造成重区分度的 计算 （4题）
总数	9题	13题	7题	11题	11题

## 二、根据测量目标确定测量的内容和形式

（一）在对测量目标进行分析之后，我们就可以根据不同情况确定测量的内容，即确定对个体的哪些行为和心理特征进行测量，并由此确定所采用的测量工具的形式。例如，用于选拔的测量工具要求对个体将来的行为进行预测，那么各种能力测验、人格测验、动机测验，以及面试、情境模拟测验都是合适的。如果只想对培训效果或者员工的工作表现和对工作的满意度进行了解，那就分别需要采用成就测验、行为观察和团体健康度评定。

（二）一般来说，测量的目标应当是明确的，测量目标决定测量内容，而测量内容又对测量工具的操作形式有所制约。例如，对成就、一般能力倾向、人格的测量，可以采用纸笔测验的形式，而对个体的言语交流能力、社交能力以及机械操作能力就不能采用纸笔测验的形式，而应采用各种操作性测量方法。

（三）确定测量内容和形式是决定性的关键，它决定测量有效与否。下面列举一个针对测量目标进行具体测量内容和测量形式设计的实例：

某商场需要招聘一批具有一定工作能力、有敬业精神并且掌握一定专业知识的摄像设备的销售人员。

第一步——确定测量目标：对应聘者未来的工作能力、工作动机进行预测，并且对应聘者现有的专业知识（对摄像的知识）进行了解。

第二步——设计测量内容：根据销售员任职资格的要求，其工作能力可以分解为以下内容：言语交流能力、说服能力、推销技巧。对成就动机的测量可以预测应试者未来工作的敬业程度。另外，通过对摄像知识的考试可以测查应试者的专业知识。

第三步——设计测量形式：根据测量内容建议采用情境模拟测验形式对工作能力进行测量，用自陈量表形式对动机进行测量，并设计一项标准化的摄像知识的纸笔考试。

### 三、根据测量的对象确定题目的形式

在编制测量工具时，还应该考虑使用测量工具的团体的组成和特点，如被试（接受测量的个体）的年龄、智力水平、受教育程度、社会经济地位、文化背景、阅读水平等因素。测量工具针对的对象不同，测量工具采取的形式、题目的编写、测量题目的难度、取样的范围都可能有所不同。例如，如果施测对象是没有阅读能力的个体，在编制测量工具时应该考虑采用操作性测量的形式，而不能采用纸笔类测验。被试群体的特点有很多，到底哪种变量更重要，取决于测量工具的目的和类型。

## 第十二章 测量工具的建立

### 第二节 题目的设计与测量的编制

#### 一、题目的来源、编写原则和分类

##### （一）测验题目的来源

##### 1. 收集题目的方法

在测量的内容及形式设计完成之后，就可以根据所设计的内容和形式要求来收集和编写题目。收集题目的方法有很多，包括从现成的测验中选取、按照现有理论设计、请专家设计等。

（1）最简单最直接的方法是，从已经出版的各种标准的测量工具中选择合适的题目。例如，编制能

力测验就可以从已出版的包含所要测量能力的测验中选取题目。当然，这样做必须注意尊重原作者的知识产权。

（2）现成的理论从来都是设计测验题目的绝好参考。如编制态度测量量表，则有关态度的类型、定义等理论都有参考和指导作用。又比如，设计职业兴趣测验，可以按照斯特朗或霍兰德等人的著名理论构架，即把职业兴趣分为六大类型，来编制相应的题目。

（3）专家是设计测验的重要资源。在实际操作上，既可以直接请专家出场设计题目，也可以参考专家的有关经验、建议或以往的工作。以人格测验为例，描述人格的术语可作为题目的来源。阿尔波特

（Allport）等人曾总结出 17953 个描述人的特点的形容词，将这些词进行归纳后，就可作为编制题目的参考。

## 2. 收集题目时应注意的问题

（1）题目的来源要尽可能地丰富，这样测量内容不至于偏颇，并能提高行为样本的代表性。

（2）题目要有普遍性，尤其是成就测验，要保证所有施测对象都学过该测验中题目所包含的内容。

（3）在编制智力或能力等本身不应体现文化影响的内容的测量题目时，要尽量避免文化背景差异的影响。也就是说，像能力测验这样的测量，应当对不同性别、种族、肤色、各种亚文化的群体具有公平性。

### （二）题目编写的原则

对测量工具题目进行编写要遵从某些一般原则，这些原则可以归纳为内容、语言、表达与理解四个方面。

#### 1. 针对题目内容的原则

（1）要求题目的内容符合测量工具的目的，避免贪多而乱出题目；

（2）内容取样要有代表性，符合测量工具计划的内容；

（3）各个试题必须彼此独立，不可互相重复或牵连，切忌一个题目的答案影响对另一个题目的回答。

#### 2. 针对题目语言的原则

（1）使用准确的当代语言，不要使用古僻艰深的词句；

(2) 文句须简明扼要，既排除与解题无关的陈述，又不要遗漏解题的必要条件；

(3) 最好一句话说明一个概念，不要使用两个或两个以上的概念；

(4) 意义必须明确，不得暧昧或含糊，尽量少使用双重否定句。

### 3. 针对题目表达的原则

(1) 尽量避免主观性和情绪化的字句；

(2) 不要伤害被试感情，避免涉及社会禁忌或隐私；

(3) 避免诱导和暗示答案；

(4) 避免令被试为难的问题（被试没有明确结论或羞于启齿的问题）。

在对人格和态度测验的题目编写过程中，有时会难以避免地涉及一些敏感性问题，如性关系、自杀倾向等。被试对这些问题的回答往往容易依照社会规范来回答，这就是回答的社会赞许性倾向的问题。社会规范告诉人们应该做什么和不应该做什么，如果被试的回答有违规范，他就会担心得不到社会赞许，甚至使他人对自己形成不好的印象，在这种担心下，被试就可能做出附和社会规范，有利于自我形象的虚假回答。为了解决这个问题，德里克·菲利普斯（Derek L. Phillips）列举了几条防止出现规范性答案的策略：

**①因为从心理上讲，否定一个答案比肯定它更为困难，所以命题者应避免具有某种行为，使他不得不在确实未有该行为时才予以否定。如，“你多久才违反一次工作规范？每天一次？每周一次？每月一次？从不？”。**

**②假定对规范无一致意见。如，“有些医生认为饮酒有害，而其他一些医生则认为有益，你认为如何？”。**

**③指出该行为不是异常的而是普遍性的，即使它可能有违规范，也是常见的。如，“多数人都有一定的离婚倾向，你呢？”。**

### 4. 针对题目理解的原则

(1) 题目应有确切答案，不应具有引起争议的可能（创造力测验、人格类测验例外）；

(2) 题目内容不要超出受测对象的知识 and 能力范围；

(3) 题目的格式不要引起误解。



### （三）题目的分类

题目的种类很多，根据被试所做出的反应（回答）方式进行分类，可以分为提供型和选择型题目。

1. 提供型题目（self-produced answer question）要求被试提供答案，如问答题、填充题。另外，各种操作性题目，如画图、表演、完成某项任务、回答问题等，都可以算作提供型题目；

2. 选择型题目（selective question）则要求被试在提供的备选答案中选择正确答案，如是非题、匹配题、选择题。

由于操作性题目往往没有什么固定模式，所以本节主要介绍纸笔类测验的题目类型。表 12—2 列出了几种常见题型。

表 12-2

几种不同形式的测量举例

**一、问答题例**

指示语：试以 2000 字说明下列问题：

1. 比较回答者与客观题目的优劣。
2. 解释进行客观测量工具项目分析的理由。

**二、填充题例**

指示语：在每个空格内填上合适的字句：

1. 客观测量工具的信效性的关键是：
2. 编制一个预测某种工作成功程度的测量工具的第一个步骤是：

**三、是非题例**

指示语：如果该题正确，请在“是”上划“/”；若该题不正确，请在“否”上划“/”。

- 是 否 1. 心理测量工具只是心理测量的方法之一。
- 是 否 2. 社会赞许性反应是一种由于被试在一项特质上评了高分，而在另一项特质上也评高分的倾向。

**四、匹配题例**

指示语：请在右边选择合适的答案的编号，填在左边每题的编号前。

- |                |         |
|----------------|---------|
| ___1. 团体智力测量工具 | a. 比奈   |
| ___2. 个体智力测量工具 | b. 达尔文  |
| ___3. 兴趣问卷     | c. 高尔顿  |
| ___4. 人格问卷     | d. 歌德斯  |
| ___5. 顿悟相关     | e. 达尔文  |
| ___6. 感知觉测量工具  | f. 罗夏克  |
|                | g. 斯皮尔曼 |
|                | h. 斯特朗  |
|                | i. 威伦斯  |

**五、选择题例**

指示语：下面哪个答案最合适？

1. “决不”、“有时”、“总是”等词语，常常会对毫无所知的被试产生暗示作用。我们称这些词语为
  - a. 普通修饰词
  - b. 压词
  - c. 反应定势
  - d. 特殊决定词
2. 某女孩，年龄八岁四个月，心理年龄为九岁，她的 IQ 是多少？
  - a. 88
  - b. 90
  - c. 113
  - d. 120

## 第十二章 测量工具的建立

### 第二节 题目的设计与测量的编制

#### 二、常用题目类型与编写要领

##### （一）问答题

##### 1. 问答题的主要优点

问答题的主要优点在于它能够测量被试组织材料的能力、综合能力和文字表达能力，有的甚至可以测量评价能力和创造能力，而这些能力是其他客观题难以测量的。相对来说，问答题有几大优势：

- （1）较好编制，题目无需太多；
- （2）不需准备备选答案，答案是由被试自己生成的；
- （3）可以避免被试随机猜测答案的可能。

## 2. 问答题存在的问题：

- （1）问答题一般回答时间长，占分数大，因此题量不宜太大，所以能测量到的内容也有限，对行为的取样受到局限。
- （2）问答题的评分标准不容易标准化。相对来说，评分者在掌握评分方法时可能有相当的主观性，而且不同评分者的评分结果很难保持高度的一致。
- （3）问答题的评分容易受书写的整洁程度、个人成见等无关因素的影响。

## 3. 编制问答题应注意的问题

一般来说，在可以用客观题施测的情况下，尽量不要采用问答题形式。如果需要采用问答题施测，题目编写者应该使问题及评分标准尽可能地客观，在编制时应注意以下几点：

- （1）问题应清楚而且明确，使被试了解答题要求；
- （2）题目的数量不要太多，以免变成速度测验；
- （3）在编制题目时应该有一个理想答案或一系列答题标准，同时对另外一些可接受的答案应有所规定和说明。

## （二）选择题

1. 选择题通常包括两个部分：一是题干，即呈现一个问题的情境，另一部分为选项，即对问题的几种可能的回答。

选择题适用于文字、数字和图形等不同性质的材料，可以考察记忆分析、鉴别推理、理解和应用知识的能力，也可以考察对某一事物的看法和观点。

## 2 选择题的主要优点包括

- (1) 适用范围广，从一般知识到复杂能力的测量均可使用；
- (2) 题意明确，被试的反应简单，容易计分；
- (3) 与其他形式的客观题相比，更少受猜测和反应定势的影响，评分较为客观；
- (4) 选择题的题量可以较大，考察的范围更广，取样代表性较高。

3. 选择题的不足之处在于诱答难以编制，诱答的数量要求多，而且还要似是而非，让那些不知道正确答案的被试感到无从选择，这相对来说较为困难。另外，通过选择题较难测出个体组织能力、表达能力和创造性等特点。

## 4. 编制选择题的一般原则

(1) 题干所提出的问题必须明确，使用简单而且清晰的用词。做到即使被试不看选项的情况下，题干本身的意义也是完整的。

- (2) 不要将选项夹在题干中间。
- (3) 选项要简练。
- (4) 除特殊情况，所有选项的长度应该大致相等，而且与题干的联系要紧密。
- (5) 避免在题目中出现帮助受测者猜测正确答案的线索。
- (6) 对于人格和态度的测量工具，题干的陈述应该不带任何倾向性。
- (7) 如果选项是数字、日期、年龄等有逻辑顺序的材料，则最好仍按顺序排列。
- (8) 答案在选项中的位置应当随机出现，没有任何规律，避免被试猜测。

## (三) 是非题

### 1. 是非题的优缺点

是非题的共同特征是只有两种可能的反应，其中一种是肯定的（同意、正确、是），另一种是否定的（不同意、错误、否）。因此是非题可以当作是只有两个选项的选择题。

是非题容易受被试反应定势和猜测的影响，测验分数的可靠性不如选择题。所谓反应定势（reaction set）

就是指部分被试在回答问题时，其答案的选择建立在题目的形式或位置上（如偏向正面回答或否定回答），而不是建立在题目内容的基础之上。

当然，是非题也有其长处。它能很快书写和阅读，因此题量可以较大，便于广泛取样；记分也比较客观。

## 2. 编写是非题时的注意事项

（1）测查的内容应以有意义的事实、概念或原理为主，不要考察被试对无关紧要的问题或琐碎的细节的辨别能力。

（2）每道题只能包括一个重要的概念，避免两个或两个以上的概念出现在同一题目中，造成“半对半错”或“似是而非”的情况。而且还要把各个概念放在题干的重要位置上。

（3）决定一个题目是否正确，应以重要的概念为基础，不要着重于对次要的或琐碎的观念的判断。

（4）除特殊情况，尽量避免否定的叙述，尤其是要避免双重否定的叙述。因为采用否定的叙述容易使人困惑，否定词也容易被一些粗心的被试所忽略。例如，题目“猜测的校正不是永远不可以使用的”就不如改为“我们有时可以校正猜测”。

（5）测验中正确的题目与错误的题目，它们的长度、复杂性应尽量一致。

（6）正确的题目数与错误的题目数应该基本相等，两种题目应按随机方式排列。

## （四）匹配题

### 1. 匹配题的优缺点

匹配题可以说是选择题的一种变式。匹配题一般包括多个反应项（匹配项）和多个刺激项（被匹配项），用反应项来匹配刺激项。匹配题有完全匹配（刺激项与反应项的数量相等）和不完全匹配（反应项目多于刺激项目）两种形式。通常，刺激项目和反应项目分别排成两列。

匹配题容易编制，而且可以在短时间内测量大量相关联的材料，覆盖面较广。但它一般只能测量简单记忆的事实材料或概念关系，并且要求编制的选项必须是同质的。

### 2. 对编制匹配题的建议

(1) 刺激项目和反应项目应该分成两列，通常反应项安排在右边；

(2) 配对数目不可过多或过少，最好使用不完全匹配，使反应项数目多于刺激项数目，并且最好不限制每个反应项被选择的次数，这样可以降低猜测的概率。

(3) 匹配题的反应项与刺激项的性质必须相近。选项如存在逻辑顺序，应按顺序排列。

(4) 应对匹配方法、匹配的依据加以明确的规定和说明，同时说明反应项可以被选择的次数。

(5) 同一组的反应项与刺激项最好印在同一页纸上，以免造成答题时间的浪费。

(6) 反应项与刺激项应以不同形式的序号加以标识，例如在反应项前冠以数字，刺激项前则冠以英文字母或甲、乙、丙、丁，以免混淆。

## (五) 填充题

### 1. 填充题的优缺点

填充题比较容易编制，不受猜测的影响，在评估专业知识方面特别适合。另外，在前面介绍的投射测验——“语句完成测验”中也用到填充题。当然，填充题也有其局限性，它不能测量更为复杂的知识和能力，题目有时也会有多个答案，计分不能完全客观。

### 2. 编制填充题的注意事项

(1) 最好采用问句形式。如果需要使用未完成句子，则填充处应尽量放在句子末尾。

(2) 使用直接问句的形式，可避免产生对题意的误解。

(3) 如果是填空形式，填充处不可太多。过多空白会使题意不明确。

(4) 每题最好只有一个答案，答案最好简短而具体，有利于评分。

## (六) 操作性测量形式

在很多情况下，操作的方法和过程是重要的测量目标，而这是纸笔测验无法测量的，这时可以采用操作性测量形式。操作测试题要求被试对未来真实情境中的行为进行模拟，其真实性要高于纸笔测验。

### 1. 操作性测量形式的分类

操作性测试题通常分为着重过程和着重结果两种形式。考察仪器操作、演讲、演奏乐器和其他各种技

艺，就需要采用着重过程的形式，这些内容需要在工作过程中进行评鉴。而有些活动则要在活动结束后考察结果，例如文章、图画等。当然，也有的需要同时考察过程和结果，如工作样本测验。

操作性测试有多种不同的分类方法，按测试情境的真实性程度可以分为：

（1）纸笔的操作性测试：虽用纸笔但偏重于模拟情境下知识的应用。如编制、编写某项操作计划、步骤、注意事项等，如公文筐测验。

（2）模拟操作测试：强调正确的程序，被试需要在模拟情境下完成与真实活动相同的动作，如驾驶员的考核等。

（3）工作样本操作测试：其真实性最高，包括了真实作业的全部要素，但是是在有控制的条件下完成的。如司机在标准场地内的考核、师范学生的教学实习等。

## 2. 设计操作性测试题的主要原则

（1）明确所要测量的目标，并将其操作化。即要进行工作分析，辨认出操作中最重要因素，找出具有代表性的工作样本；

（2）要建立作业标准，规定通过此项作业的最低标准。

（3）选择合适的真实性程度。通常情况下，真实性程度越高，模拟的代价越大。

（4）指示语简单明确，让被试知道要干什么和在什么条件下去做。

（5）有明确的计分方法。

有些项目根据完成题目的数量和错误次数客观计分就可以了，有些项目的评分则较为困难，这种项目可以采用“作品量表”来计分。作品量表（standard sample scale）一般包括一系列按顺序排列的不同作业程度、水平、质量的标准样本，评分时参照这些标准样本对被试结果进行评分。

## （七）面试

### 1. 面试的优缺点

### 2. 编制面试题目的原则

对于结构化面试，其面试问题都是预先确定的。对问题的编制要注意问题的明确性，同时，对于需要

掌握细节的问题，一定要设计若干追问问题。其他原则与问答题的编制类似。

## 第十二章 测量工具的建立

### 第三节 题目的编排

#### 一、题目的选择

##### （一）对题目的初步选择

对题目进行初步选择的主要依据是：题目本身的性质、对预期测量目标的可测量程度。这一步可以细分为如下几个步骤：

1. 检查题目是否符合细目中某一单元格内容的要求。
2. 根据细目表对各部分所要求的比例选择适当数量的测试题，使测量工具尽量覆盖整个细目表的内容。
3. 检查题目是否叙述明确清楚，是否提供了额外线索。
4. 检查题目是否适合将要施测的对象和施测的条件。
5. 检查题目的难度是否恰当，一般来说，难度定为 0.50（50%的人答对此题）较为合适（对能力、技能类的题目要求除外）。
6. 检查所选择的题目是否彼此独立，没有重叠，即回答某一问题所需的知识是否与回答其他问题无关。

##### （二）在选择题目时还要注意测量工具的长度

在一个测量工具里究竟包括多少题目比较合适，既要看是否能完全满足测量目的的要求，也要根据测量工具的时间限制和被试的年龄和阅读水平而定。通常应该在正式测试前，挑选几个被试对测量工具进行预测试，以观测测量工具所需时间。并对测量工具长度进行调整。

#### 二、题目的编排

在对测试题初次选择完毕之后，就应决定如何对选出的题目进行最佳编排。测试题的编排方式因测量



工具的类型不同而有所差别，下面是题目编排的一般原则：

（一）将测量相同因素的测试题排列在一起。

（二）尽可能地将同一类型的测试题组合在一起。这样只需对每一类型的试题作一次说明，也方便被试的回答，同时还可以简化记分工作和对题目的统计分析。

（三）难度测验的题目应按由易到难排列，这样可以鼓舞被试的士气，避免某些被试一开始就因较多题目回答不出而失去信心。这种安排也可以使被试熟悉反应程序，解除紧张情绪，同时还可避免被试在难题上耽搁较长时间而影响了后面的回答。

（四）对于人格测验，应尽量避免将测量同一特质的题目编排一起，防止被试猜测出题目所要测量的因素。

### 三、测试题的预测试和分析

（一）题目分析的含义

测试题编排完成并不意味着这项测量工具的编制就此完成。前面对题目的选取只是依靠编写者的主观经验，题目的效果如何还需要进行定量的客观分析。这时的测试题还只能叫作预备测试题，还需要获取被试对这些题目的反应的材料，为进一步筛选题目和为编排测量工具提供客观依据。也就是说，必须将预备测试题对一定规模的小样本被试进行施测，获得数据以进行校验、修订。这一过程称为测试题的预测试(pilot test)。数据收集上来之后，就应该利用这些数据对题目进行分析，删除不好的题目或对题目进行修改，这个过程称为题目分析(thematic analysis)，具体内容包括对题目的项目分析和对测验信度、效度指标的检查。有时候，题目需要经过多次预测试和分析，不断修改后才能达到测量学的要求。

当预测试完成后，可以根据预测结果进行题目分析，对每个题目的具体分析称为项目分析(item analysis)，主要是指根据题目的难度、区分度、备选答案的合适度等数量指标来对题目进行分析。根据分析结果，再对题目进行选择，最终编制出较好的测量工具。

（二）测试题的预测试应注意的问题

1. 预测试对象必须和将来正式测试的对象相似，取样应注意其代表性，人数不必太多，但不能太少，

一般不应少于 30 人。

2. 预测试的实施过程和情境应力求与将来正式测量工具实施时的情况相似。

3. 预测试的时限可稍宽些，最好使每个被试都能将题目答完，以便搜集充分的反应资料，使统计分析的结果更为可靠。

4. 在预测试过程中，应将被试的各种反应情况随时加以记录。如，记录在不同时限内一般被试所完成的题目数、题意不清之处、被试的态度等，以便在修改测验时作为参考。

#### 四、测验使用手册

任何标准化的测验都必须提供给用户相应的说明书，即测验手册。测验手册对如何使用该测验进行说明，是测验使用说明书，同时也是用户评价、比较测验优劣的重要依据。

测验手册的内容一般包括：

（一）测验的一般信息：标题名称、作者、施测形式、记录形式。

（二）测验目的和作用。通常手册应指出测验可作何种用途（如选拔、指导、评价），适用对象，以及所测量的内容（如测量何种能力、特质）等。

（三）测验编制的理论背景和题目的来源。一些手册还提供选择题目的统计指标。

（四）测验的实施方法、时间要求及注意事项，包括对施测现场环境的要求、指示语和对主试的训练要求。

（五）测验的标准答案和记分方法。

（六）常模表、常模适用团体及测验计分系统、解释系统，以及样本的选择方法。

（七）测验的信度资料和效度资料，以及这些资料取得的条件和情境，包括调查的样本和时间。

测验使用手册可以说是测验的“门面”，其编写的好坏、形式十分重要。一个测验的编制、修订可能花费巨大的时间、精力、财力，但最后的主要表现形式却很简单，往往只是一个使用手册和题本，使用者往往只是从这些形式上了解测验的价值。因此，编写使用手册一定要到位，充分体现测验的功能和意义。

## 第十三章 测量的实施与计分

### 第一节 实施测量操作的要领与误差控制

#### 一、测量的操作程序

##### （一）标准化指示语

##### 1. 指导语的概念

在施测过程中应该使用统一的指示语。指示语（instruction）是在测量实施时说明测量进行方式以及如何回答问题的指导性语言。指示语通常有两种：一种是对被试的，另一种是给主试的。前者应该力求清晰和简单，向被试说明应该做什么，即如何对题目作出反应。这种指示语一般印在测验的开头部分，可以让受测者自己阅读，同时又在测验开始前由主试口头说明，以确保被试确实掌握要求。

##### 2. 指导语的内容

一般来说，对被试的指示语包括：

- （1）如何选择反应形式（画圈、画勾、填数字、口答、书写等）；
- （2）如何记录这些反应（答卷纸、录音、录像等）；
- （3）时间限制；
- （4）如果不能确定正确反应时，该如何去做（是否允许猜测等）以及计分的方法。
- （5）例题。当题目形式比较生疏时，应该给出附有正确答案的例题；
- （6）某些情况下告知被试测验目的。

##### 3. 指导语的要求

- （1）测验的指示语必须清楚、明确、易懂、有礼貌。
- （2）标准指示语通常要求主试在指示语念完后，要询问被试有何问题。

##### （二）测验时限

大多数典型作为测验是不受时间限制的。例如，人格测验中，被试的反应速度并不很重要，所以

人格测验一般不限定被试完成测验的时间。但在最高作为测验中，速度是需要考虑的重要因素之一。

大多数测验既要考察反应的速度，也要考察解决有较大难度题目的能力。那么，测验时间定为多长比较合适呢？通常，在能力和成就测验中所使用的时限，以大约 90%的被试能在规定时间内完成测验为标准。如果题目从易到难排列，力求使大多数被试能在规定时间内完成他会答的题目。确定测验的标准时间一般采取尝试法，即通过预测试来确定。

### （三）测验的环境条件

#### 二、主试的职责

主试（tester），俗称考官或主考人，是控制测试进程的主要人员。主试的经验如何往往会对测验的结果产生影响。对于个体测验、投射类测验、操作性测试和面试，主试的作用就更为重要。测验手册中一般有对主试职责的详细说明，主试应严格按照此说明主持施测。通常，需要对第一次主持一项测量的主试人员进行培训。一般来说，主试应做以下工作：

#### （一）测验前的准备工作

1. 预告测验
2. 熟悉测验指示语
3. 准备测验材料
4. 熟悉测验的具体程序
5. 确保满意合适的测验环境

#### （二）测验中主试的职责

1. 主试的主要职责是按照指示语的要求实施测验，在被试询问指示语意义时，作进一步澄清，但注意不要作任何暗示。另外，在测验时，主试还要注意不要讲与测验无关的话，并能够对测验中的特殊情况做出灵活的解决。

2. 一般来说，主试应做测试记录，记录下测试现场发生的任何细节，因为它们可能和结果评价、解释有关，这对那些不用录音录像设备记录的测验来说，是很有帮助的。此外，这些信息还可为今后

修订测验提供一定依据。

### （三）建立协调的关系

建立协调关系就是要求注意必须促使被试尽可能地对测验感兴趣，遵从指示语，认真合作地进行测验。因此无论在个体或团体测验中，主试都应该采取热情、友好并且客观的态度，这是建立协调关系的前提。

## 三、测量实施过程中可能导致误差的各种影响因素

### （一）主试对测量结果的影响

测量结果往往会受到主试的各方面因素的影响：

#### 1. 主试的人格特点

主试的不同特点对测验的实施及测验的评分等环节都有影响。有些主试可能自己就不大善于建立和处理人际关系，对他来说，在测验实施过程中与被试建立协调关系较为困难，因而由他施测的被试的测验结果可能就会受到影响。有些竞争性很强的主试，在测验时也往往苛求受测者。而有些主试过于宽容随和，在测验中给予过多的关心甚至评以高分，也会使测验出现偏差。

#### 2. 主试的期望

在有些情况下，实验者所获得的资料及实验结果会受其本身期望的影响，这种现象称为罗森塔尔效应（Rosenthal effect）。

在心理测验中也同样存在这种效应的影响。例如，要求正在进行智力测验实习的研究生给测验中一些含糊、不清楚的答案记分。将评分的研究生随机分为两组，告诉其中一组他们所评分的答案是聪明的被试回答的，而告诉另一组研究生，他们所评判的答案是由较笨的被试回答的。结果发现，在对同一答案进行评分时，被告知答案由聪明被试做出的这组研究生所评分数高于另一组。

当然，相对来说，主试对测验结果的影响仍是有限的，是可以通过一定方法有效克服的。具体来说，就是要力求做到测验实施过程的标准化，将主试的个人因素对测验结果的影响尽可能降到最低。

### （二）被试特点对测量结果的影响

## 1. 测验的技巧与练习因素

(1) 测验的技巧。显然，如果某个被试熟悉测验程序及题目形式，而另一名被试是面对全然陌生的测验材料，这两者的测验结果是无法比较的。具有某种测验技巧的被试，他们能够觉察正确答案与错误答案的细微差别，知道合理分配时间以及适应测验形式等。通过应用这些技巧，他们通常比那些与他们能力相等但是测验技巧较差的被试获得更高的测验分数。因此，在测验标准化时，应尽量设法使每个被试对测验材料的步骤和所需技巧有相同的熟悉程度。必要时，可以增加练习测验，使所有应试者同等程度地熟悉测验形式。

(2) 练习效应。有不少研究发现，应试者参加相同或重复的测验，会由于练习效应而使测验成绩提高。

## 2. 焦虑和动机因素

(1) 应试动机。被试参加测验的动机不同，会影响其回答问题的态度、注意力、持久性以及反应速度等，从而影响最后的测量结果。在测量成就、智力和能力倾向等内容时，如果被试动机不强烈，就不会尽力回答，导致对被试能力的低估。

(2) 测验焦虑。焦虑 (anxiety) 是一种不愉快的、表现为焦急、恐惧和紧张的情绪体验，它主要是由于对可能出现的结果的担心或对应付某一局面的能力的担心而造成。大多数人都在测验前和测验中感到焦虑，故又称测验焦虑或考试焦虑 (test anxiety)。测验焦虑通常会影响到测验的结果。一般来说，适度的焦虑会使人的兴奋性提高，注意力增强，反应速度加快，从而对智力和学术性能倾向有积极影响。过度的焦虑则会使工作能力降低，注意力分散，思维变的狭窄、刻板。毫无焦虑，则往往源于对测验的动机不高，因而成绩大多偏低。因此，在测量过程中，并不必担心应试者有适度的焦虑水平，但应注意消除可能造成应试者过于紧张的外在因素。

## 3. 反应定势

反应定势也称为反应的方式或反应风格，简单地说，就是每个人回答问题的习惯方式。由于每个人回答问题的习惯不同，可能会使有相同能力的被试获得不同的分数。影响测量结果的反应定势主要

有以下几种：

- (1) 求“快”与求“精确”的反应定势。
- (2) 偏好正面叙述的反应定势。
- (3) 偏好特殊位置的反应定势。
- (4) 偏好较长选项的反应定势。
- (5) 猜测的反应定势。

#### 四、猜测的校正

##### (一) 猜测修正的方法

常用的猜测修正公式为：

$$s = R - \frac{W}{n-1} \quad (13.1)$$

其中， $s$  是正确分数， $R$  为被试答对的题目数， $W$  为被试答错的题目数， $n$  为选项数目。

使用公式 13.1 时，必须分别算出答对及答错的题数，要特别注意不可将未答的题数归并到答错的

题数中去。 $n$  的大小视选项的数目确定。若为是非题， $n=2$ ，则  $s = R - \frac{W}{2-1} = R - W$ ；若为三择一选择题，  
则  $s = R - \frac{W}{2}$ 。

##### (二) 猜测修正的优缺点

###### 1. 猜测修正的优点

- (1) 修正公式可避免减低测验的信度。因为如果强调倒扣分，被试就不敢盲目猜测。
- (2) 修正公式可以反映被试真正的能力和水平。
- (3) 使用修正公式对那些不能答完全部试题的被试来说比较公平。

###### 2. 猜测修正的缺点

- (1) 公式假设不成立，因为被试答错试题，并非完全瞎猜。大多数情况下，均是先舍弃部分诱答，再就剩下的几个选项来猜测。有时被试答错试题，是因为观念模糊、记忆不清或者粗心大意，并非乱

猜。

(2) 只要所有被试能答完全部题目，或者是未答的题目数相同，则猜测修正无实质作用。此种情形下其相对分数（如  $Z$  分数、 $T$  分数或百分等级）完全相同，因为依据心理计量学原理，校正后的分数虽与校正前不同，但其相关系数为<sup>1.0</sup>。为此，这两种分数对于决定分数的高低具有相同的作用。采用猜测修正，徒然增加计分的复杂性。

(3) 未采用猜测修正对信度并无重大影响。

(4) 实际生活中，常常需要仅凭借部分知识来作判断，因此完全不许猜测也与实际生活不符。

3. 对于是否需要采用猜测修正并无定论。但是，如果是需要速度的测验，也就是说，有被试答不完所有试题时，可以采用猜测修正。如不采用猜测修正，则在编制测题时要注意：

(1) 选项数目要多，4~5 项较合适；

(2) 题目数量加多；

(3) 时间要充裕；

(4) 经常以项目分析评鉴试题，并淘汰不好的试题。

## 第十三章 测量的实施与计分

### 第二节 实施测量记分的要求与误差控制

#### 一、计分的基本步骤和要求

计分的基本步骤主要有三步：

##### (一) 记录反应

及时和清楚地记录被试的反应。如果是纸笔类测验，被试的答案将由被试自己记录在答卷上。如果是口头回答、操作演示回答等，则需要主试进行记录。这种情况下，可以用录音和录像等较为技术化的记录方法，以避免记录时记忆的困难和记忆错误。

##### (二) 检索标准答案



标准答案有时又称计分键。选择题测验的计分键是每一道题的正确答案的号码或编排字母；填空题的计分键是一系列正确答案以及所允许的变化；问答题的计分键为各种可接受的答案的要点；操作题的计分键则是指具有某种特点或能力的个体的典型反应。如果以上反应需要加权，权数也应在计分键中标明。

### （三）反应和标准答案的比较

也就是将反应归类或赋予分数值。客观题的程序很清楚，但当评分者的判断可能成为影响分数的一个因素时（如问答题），就需要对评分的规则作详细的说明。评分时将个人的反应和评分说明书上所提供的样例进行比较，按最接近的答案样例给分。

## 二、问答题计分

（一）问答题的主要缺点就是评分不够客观，也就是说，问答题的计分经常受到评分者的情感、态度的影响。问答题中常见的误差有：宽容定势和晕轮效应。宽容定势（leniency set）指主试的计分过于宽松，即使没有回答出题目所要求的答案，评分者也给予较高的分数；晕轮效应（halo effect）指给予被试某道题较高分数仅仅是由于被试在另外一些试题上获得了高分，也就是说对被试的一般印象影响到具体某个问题的评价。

（二）为了使问答题的计分更加客观和可信，主试应该首先考虑采用何种计分程序，整体计分还是分析计分。整体计分（global scoring）就是评分者根据总体印象给答案评一个总分。整体计分在实际中应用较为普遍。分析计分（analytic scoring）是给问答题的不同部分分派不同的权数，按照各部分的要求对答案中所包括的信息和技能评分，最后将各部分的权数和得分组合起来得到该问答题的分数。分析计分往往有答题的详细标准，因此相对更为客观。

### （三）问答题计分的一般原则：

1. 与测量目标无关的回答不予计分，或单独给分数。评分应依据被试对问题的回答是否充分和恰当，所有其他因素，诸如文风、答案长短、书写等，在计时时应尽量不予考虑。
2. 确定标准答案。问答题应具备一定的标准答案和评分标准。

3. 评分时最好按题目顺序而不是按被试顺序进行，即对所有被试第一个问题答案计分完毕之后，再给下一题的答案计分。这样可使计分标准一致，亦可避免“晕轮效应”的影响。

4. 最好在评阅时不知道被试的名字，以减少个人偏见。

5. 安排两个或两个以上的主试来给问答题计分，取其平均值作为被试的分数。也可由一人在第一次评阅后，再作第二次审查，以确定评分是否偏颇。

三、客观题计分

客观题的一个主要优点就是计分简单、客观。客观题的分数可由一个一般的工作人员利用计分套板和计分器很快地、准确地算出。客观题的计分由题目的形式决定。能力测验和成就测验中，通常是按正确答案给 1 分，不正确计 0 分来统计。例如，一个有 50 道选择题或是非题的测验，其分数的分布将是 0—50 分。在人格测验中，没有答案正确与否的区分，但每种反映特定倾向的选项都可以用一个数字或符号进行标定，最后统计被试选择这种选项的次数。有时，不同的反应依据主试认为的重要性不同也可以给予不同的权数。另外，我们还可以根据被试回答问题时的确定程度给予不同的权数。表 13—1 表示这种“信心权数”的应用。

表 13—1 伊贝尔是非题加权方法

被试对题目确定程度	假定答案为“是”时得分	假定答案为“否”时得分
是	2 分	2 分
不可能	1 分	0 分
不知道	0.5 分	0.5 分
可能不是	0 分	1 分
不是	-2 分	2 分

第十四章 人事测量结果的解释

第一节 常模与基于常模的解释

一、常模的选择

### （一）常模样本的构成

1. 常模样本通常是具有某种共同特征的人所组成的一个群体，或是该群体的一个样本。对人事测量来说，常模样本往往是该测验的适用群体（即测验总体或常模总体）的一个样本。例如，建立一个管理者管理技能测验，它的常模样本就是管理者。如果是区分高层管理者与中层管理者的管理技能测验，那么常模样本就分别是高层管理人员和中层管理人员。又例如，如果是建立一个营销人员销售技能的测验，那么常模样本就是营销人员。或者，如果是一项专门用于考察应聘大学毕业生一般能力素质的测验，常模样本就是大学生。

2. 因为有的测验可以适用于多个不同类型的群体，而应试者的分数需要与同类型的群体进行比较，所以有时一个测验可以有很多常模样本。例如，很多能力倾向测验都分别有中学生的常模、大学生的常模等，与不同的常模样本比较，同一分数得出的相对等级不同。因此在做常模参照分数的解释时，必须考虑常模样本的组成。

3. 在编制测验时，常模的选择主要依赖于对测验将要施测的总体（测验适用的所有对象）的认识，常模样本往往是常模总体的一个有代表性的样本。确定常模样本的过程包括：确定一般总体、确定目标总体、确定常模样本这三个步骤。例如，编制一项适用于大学生的一般能力倾向测验，其一般总体就是大学生；目标总体是测验计划实施的对象，如某些城市的部分大学的学生；而常模样本的选取必须根据总体的性质（如性别、年龄、专业、家庭背景等）选取能够反映总体性质的代表性样本。

4. 对于测验的使用者，常模样本选择的意义在于确定一个合适的常模作为比较的标准。标准化测验通常提供许多原始分数与各种常模样本的比较转换表，解释分数时将被试的分数与合适的常模比较。有了这种比较转换表，所有的计算、换算工作都已经事先完成，就可以把通过计分得到的原始分数直接在表中查出其意义，省去了解释人员的许多麻烦。例如，在进行人员安置时，同一个测验分数就可与各种不同工种的常模进行比较，从而确定被安置人员更适合哪一工种。如何选择合适的常模是测验的使用者应注意的重要问题。如果将中学生的能力分数与大学生的常模进行比较，这样的比较结果是毫无意义的。同理，如果拿一般大学生的常模去衡量企业高层管理人员，也是不恰当的。

5. 在最终确定常模样本的结构时，有许多可能的影响因素是需要考虑的。测验性质的不同，文化教育水平不同，性别不同，都可能影响常模。例如，成就测验和能力倾向测验中，常模样本应包括目前的和潜在的竞争者；一般的能力测验的常模样本由同样年龄或同样教育水平的被试构成；人格测验的常模样本通常是同年龄段的正常成年人，有时区分为正常男性或正常女性这两个团体。其他一些因素如职业、社会经济地位、种族等，也都可能作为定义常模样本的标准。

## （二）常模样本的条件

在确定和选择常模样本时，要注意下列要求：

### 1. 常模样本的构成必须明确

一个测验可能有许多常模样本。例如，工作动机测验可以有不同性别的常模样本，不同职位层次（中层与高层管理人员）的常模样本，不同年龄段的常模样本。如果每个常模样本没有明确的界定，测验的使用者就无法确定将应试者的成绩放在哪个常模中进行比较。

在选取常模样本时，首先要保证常模样本的所有成员是同质的，可以相互比较。另外，要在测验手册中对常模样本进行明确的说明，包括常模样本的构成和特性等。如常模样本是某种特定行业或职位的人员时，应明确说明是哪种行业或职位、被试工作经验、教育背景等相关资料。

### 2. 常模样本必须是所测群体的代表性样本

常模样本应能够代表测验总体，常模样本的分数的分布应尽量与总体分数的分布相吻合。这就要求常模样本的构成与测验总体相同。因为只有这样，在我们基于少数特定数据预测、推论一般性时，才可能比较准确。如某项测验的总体是全国大学生，那么构成常模样本的大学生最好来自各个年级、各个地区、各类学校、各种学科，并且各种类型的大学生的比例最好与全国大学生的比例接近。比如，假设全国大学生男女生的比例为 2:1，那么常模样本中的男女大学生的比例也应该是 2:1。

### 3. 样本大小要适当

样本的大小要满足能够反映总体分数的分布这一基本要求。从经济的角度出发，样本无需太大，只要能反映总体分数的分布即可。

常模大小（即取样大小）由以下三方面决定：

（1）常模样本大小决定于总体的规模。总体规模小，如只有几十个人，则常模样本应包括所有成员。如果总体数目大，相应的样本也应较大，一般最低不应少于 30 个或 100 个。

（2）常模样本大小决定于总体性质。如果总体只有单一性质，则任何数目的样本都能够反映总体性质。总体性质越复杂，越需要较大数量的样本。

（3）常模样本大小决定于施测结果。根据统计学原理，样本的标准误是以样本数的平方根去除标准差。标准误越大，预测的可靠性越差。设  $S$  为标准差， $n$  为所抽样本数， $S_x$  为样本标准误，则可得：

$$S_x = \frac{S}{\sqrt{n}} \quad (14.1)$$

移项可得

$$n = \left( \frac{S}{S_x} \right)^2 \quad (14.1')$$

从公式 14.1 可以看到，样本数  $n$  越大，标准误  $S_x$  就越小，效果就越理想。假如我们从试测结果或从过去已有的研究中得到总体的标准差，同时希望提高样本的准确度，也就是说减小样本的标准误，则样本数目应随之扩大。例如，某群体标准误为 4，我们希望降到 2，则样本数目应扩大 4 倍，如果希望降到 1，则样本数目应扩大 16 倍。

#### 4. 注意常模的时效性

由于常模总体可能会随时间而改变，所以还要对常模制定的时间加以考虑，也就是说，一个常模到底能在多长时间内有效。许多人在考虑常模的合适性时，往往忽略对常模时间性的要求。由于几年前所编制的常模可能并不适合于现在的要求，因此常模必须定期修订。在选择合适常模时，应注意选择较为新近的常模。

#### （三）取样的方法

取样（sampling）是指从目标人群中选择有代表性样本的过程。从统计角度看，取样的方法有随机抽样和非随机抽样两种。前者是根据随机原则进行，而后者则没有随机性。

在确定常模时，常用的取样方法有：

### 1. 简单随机抽样

按照随机顺序表选择被试作为样本，或者是将抽样范围中的每个人或者每个抽样单位编号，随机选择，以避免由于标记、姓名或者其他社会赞许性偏见造成抽样误差。在简单随机抽样中，每个人或抽样单位都有相同的机会被抽取作为常模中的一部分。

### 2. 系统抽样

系统抽样的方法是这样：假设总体数目为  $N$ ，若要选择  $K$  分之一的被试作为样本，则可以把所有的人  $N$  分为  $N/K$  组；每个组选一个人，则刚好组成  $1/K$  的样本。或者把所有的人从 1 到  $N$  按序编号，把所有编号是  $K$  的倍数的人抽取出来，即可组成所需样本。例如，某个总体为 100 人，若抽取二分之一， $K$  为 2，样本数为 50 人；若抽取五分之一， $K$  是 5，样本数是 20。假设现在要抽取  $1/4$  的人作为样本，则  $K = 4$ ，应抽取人数为  $N/K = 100/4 = 25$ ，则将总体分为 25 组，每组抽取 1 人即可。或者把所有的人按 1 到 100 编号，凡是 4 的倍数的序号抽取出来即可。

需要注意的是，在进行系统抽样时，分组和从组中抽取，或者是编号，都必须是随机的。如果能够找到任何没有偏见的排列顺序，就可采取系统抽样方法。如果发现排列有某种内部循环规律存在，就不能如此进行。

### 3. 分组抽样

在总体数目较大，无法编号，并且总体成员又具有多样性的情况下，可以先将群体分为一定的小组，再从小组内随机抽样。例如，将全国大学生按地区分为许多小组，然后在各小组中选取一定比例的大学生作为被试。当然，这里保证抽样可靠性的关键是选定和划分小组。如果这个过程无法或难以保证随机性，则最后抽取的样本的代表性会有损失。不过，这种损失也是相对的，在多数情况下是可以接受的。比如，如果由样本代表性造成的预测误差远小于测量工具本身的计量误差，则往往可以忽

略不计。

#### 4. 分层抽样

在确定常模样本时，最常用的是分层抽样（stratified sampling）方法。它是先将目标群体的某一种变量（如年龄）分成若干层次（如不同年龄段：20—29 岁，30—39 岁，……），如  $R$  个层次，再从各层次中随机抽取若干个案。各层次的个案总和即为样本个案数目：

$$N = \sum N_i = N_1 + N_2 + \dots + N_R$$

分层抽样能够避免简单随机抽样中样本集中于某种特性或缺少某种特性的现象；它使各层次差异显著，同层次保持一致，增加了样本的代表性。使用分层抽样方法获得的常模在解释测验分数时更为实用和有效。

分层抽样也分为两种方法：

（1）分层比例抽样法。如果各层抽取的个案数目  $n_i$  是根据各层的个案数目  $N_i$  占总体数目  $N$  的比例而决定的，则

$$n_i = \frac{N_i}{N} \times n \quad (14.3)$$

其中， $n$  为欲抽取个案总数； $\frac{N_i}{N}$  代表比例，用  $W_i$  表示，上式可改成：

$$n_i = W_i \times n \quad (14.4)$$

例如，假设全国大学生 19~20 岁年龄的学生占 25%，那么我们在抽取 1000 名大学生为样本时，其中抽取的 19—20 岁的人数就应为 250 人。

（2）分层非比例抽样。当各个层次的差异很大时，就不宜采用比例抽样。因为有些层次的重要性大于其他层次，这时应该采用非比例抽样方法。这种方法的目的在于减低各层的标准差，使总体平均数的估计较为准确。应用非比例抽样法，各层样本的数目可由以下公式求得：

$$n_i = n \times \frac{M_i \cdot S_i}{\sum M_i \cdot S_i} \quad (14.5)$$

其中， $n_i$  为各层应抽取的个案数； $n$  为样本个案数； $M_i$  为各层个案数； $S_i$  为各层调查单位的标准差。

## 第十四章 人事测量结果的解释

### 第一节 常模与基于常模的解释

#### 二、常模的类型及其解释

常用的常模有发展常模、百分位常模和标准分数常模。下面我们着重介绍百分位常模和标准分数常模。

##### （一）百分位常模

百分位常模包括百分等级、四分位数和十分位数。

##### 1. 百分等级

（1）百分等级是应用最广的表示测验分数的方法。百分等级的概念很容易理解。它是指把一个总体的所有分数按大小顺序排列后，把所有分数按个数等分为 100 等份，这每一个等份对应的百分数就是这个分数分布的百分等级（percentile rank），与各个百分等级相对的分数值则叫百分位数。有了百分等级，对分数进行比较就十分方便直观，因为一个分数的百分等级实际上就是指在常模样本中低于这个分数的人数百分比。比如，如果一个原始分对应的百分等级为 98，就表示在常模样本中有 98% 的人的得分比这个分数要低。这样一来，这个分数的水平高低的意义就可一目了然。换句话说，百分等级是以百分率的形式来表示一个人的相对等级，即我们将常模样本分成 100 等份时这个人所占的等级。

（2）百分等级是一个非常有用的量值。不管样本分数分布是什么形态，是随机的还是非随机的，我们总可以给所有分数进行排列并求出所有原始分数对应的百分等级，从而可以进行分数的比较。比如，某企业对所有 400 名职员进行了基本技能考核，得出了 400 个不同的考核分数。从考核分数本身很难看得出水平高低。比如，一个人得分为 73 分。这是什么意思？满分是 100 分和满分是 80 分时，73



分的含义不同；员工中最高分为 99 分和最高分为 79 分时，73 分的含义不同；即使最高分是 79 分，但 73~79 分之间有 200 人和只有 20 人时，73 分的含义也不同。也就是说，除非我们知道 73 分相对于其他分数的位置，才能作出有意义的评价。这时，如果我们求出各分数对应的百分等级，就可以直接明了每一个分数相对于其他分数的位置了。

(3) 百分等级的计算关键在于确定在常模样本中分数低于某一特别分数的人数比例，这可以分两种情况：

①对没有分组资料的数据分布求百分等级，公式为：

$$PR = 100 - \frac{100 \times (R - 0.5)}{N} \quad (14.6)$$

其中， $R$  为排名顺序， $N$  为总人数。例如小明在 30 名同学中得第 5 名（考 80 分），则其百分等级为：

$$PR = 100 - \frac{100 \times (5 - 0.5)}{30}$$

百分等级 85 的含义是，在同学中，得分低于小明的有 85%。

②对有分组资料的数据求百分等级。先来看表 14—1 所给出的分组资料，这是把所有分数按 5 分的组距进行分组的结果。

表 14—1 某测验分数分布的中点百分等级和标准分 (N=250)

分数间距	中点	$f_f$	$C_f$	PR	$s$	$Z$	$Z_n$	$T$
74.5—79.5	77.5	3	247	99.5 (100)	2.99	76	2.51	75
69.5—74.5	72.5	11	236	97.0 (97)	2.3	10	1.83	—68
64.5—69.5	67.5	18	218	91.5 (92)	1.48	65	1.33	63
59.5—64.5	62.5	27	191	82.9 (83)	0.92	59	0.91	59
54.5—59.5	57.5	40	142	66.4 (64)	0.37	54	0.43	54
49.5—54.5	52.5	65	77	46.4 (46)	—0.19	46	—0.16	46
44.5—49.5	47.5	38	39	24.7 (25)	—0.74	43	—0.73	43
39.5—44.5	42.5	25	14	11.6 (12)	—1.30	37	—1.25	37
34.5—39.5	37.5	13	1	3.5 (4)	—1.84	31	—1.88	31
29.5—34.5	32.5	1	0	0.2 (0)	—2.41	26	—2.88	21

对这类资料中任一个分数计算百分等级的公式如下：

$$PR = \frac{100}{N} \left[ \frac{(x-l)f_p}{h} + c_f \right] \quad (14.6')$$

其中， $x$  为任意原始分数， $l$  为该原始分数所在组的精确下限， $f_p$  为该分数所在组的次数， $c_f$  为  $l$  以下的累积次数， $h$  为组距。

例如，在表 14—1 中，欲求 72.5 这一分数的百分等级，可将相应数据代入公式：

$$PR = \frac{100}{250} \left[ \frac{(72.5 - 69.5) \times 11}{5} + 236 \right] = 97.04$$

## 2. 百分点或百分位数

在决定百分等级时，我们是求低于某一测验分数的人数比例，然而有时我们却想知道位于某一比例的分数是多少。例如我们想挑选得分高的 15% 的被试，我们就必须求出相当于 85 百分等级的测验分数。在分数量表上，相对于某一百分等级的分数点就叫百分点（percentile point）或百分位数。实际上，求百分点就好比是求百分等级的逆运算。

3. 四分位数和十分位数。四分位数与十分位数和百分位数含义相似。百分位数将量表分成 100 等份。然而在许多情况下，并不需要如此精密的区分，只要分成少数的段落区间就足够了。四分法将数据分布分成四等份，实际上是第 25、50、75 等百分点分段，因而计算四分位与计算第 25、50、75 的百分点是相同的。同样，十分位的计算则与计算第 10、20、……、90 等百分点相同。

依照惯例，最低的 1/4（第 1—25 的百分等级）为第一个四分位，最低的 1/10 则为第一个十分位，依次类推。

## （二）标准分数常模

标准分是一种有相等单位的分数。这里我们介绍由标准分如何构建常模。标准分数常模有好几种，每一种都是基于特定的根据平均数和标准差转换原始分数而计算标准分的方法。常见的标准分数有： $z$  分数、 $T$  分数、离差智商（IQ）、 $T$  分数、标准九分等。由原始分数转换的标准分数，又分为线性转换

和非线性转换两种：

1. 线性转换的标准分数  $z$  分数 (z score) 为最典型的线性转换的标准分数，它是指以标准差为单位所表示的原始分数与平均数的差距。根据这个定义，可表示为下列公式：

$$z = \frac{x - M}{S} \quad (14.7)$$

其中， $x$  为原始分数， $M$  为平均分数， $S$  为标准差。由此可见， $z$  分数可以用来表示某一分数偏离平均数多少个标准差，偏离方向如何。 $z$  分数具有以下几个性质：

(1) 它的平均数为零，标准差为 1；

(2)  $z$  分数的绝对值表示这个原始分数与平均数的距离，正、负符号表示原始分数在平均数之上或之下；

(3)  $z$  分数的分布形状与原始分数相同，因为是按照线性关系将原始分数转换成标准分数的。假如原始分数的分布有偏斜，则  $z$  分数的分布也一样；如果原始分数按正态分布，则  $z$  分数的范围大致是  $\pm 3.00S$ ，即正、负 3 个标准差的范围。

(4) 由于标准分数  $z$  是以间隔量表来表示的，所以可作一般的代数运算。

由于  $z$  分数中会出现小数点和负值，而且单位过大，所以通常又将  $z$  分数转换成  $Z$  量表分数，转换方法是：

$$Z = A + Bz \quad (14.8)$$

其中， $Z$  为转换后的标准分数， $A$ 、 $B$  为常数。由于加上或乘以一个常数并不改变量表中的比较关系，所以  $Z$  分数与  $z$  分数是同质的。

如果令  $A=100$ ， $B=10$ ，那么它的平均数为 100，标准差为 10：

$$Z = 100 + 10z$$

不难看出，IQ 分数实质上就是一种  $Z$  分数，其平均分为 100，标准差为 15。

## 2. 正态化的标准分数

(1) 线性转换的标准分数尽管具有不同的平均数和标准差，但仍与原分布的形状一致。若原始分数为正态分布，则标准分数亦然。在正态情况下，标准分数还可转换成百分等级。这种转化可利用正态曲线的面积表。因为在正态分布中，标准分数（如  $z$  分数）与曲线下的面积（在任意两个分数点间的人数所占比例）有特殊的固定的数学关系。

当原始分数不呈正态分布时，也可以进行面积转化，而使分数转成为正态分布。由这种方式所得到的分数就叫正态化的标准分数。为了使分数正态化，必须有某种依据能假定所测特性的分数事实上应该是呈正态分布的。假如无法作此假定，强行使之成正态分布，只会扭曲分数分布。因此，只有当所得的分数趋近于正态而只是由于取样误差使其稍有偏异时，才可计算正态化的标准分数。这种情况在对大量而且异质的样本进行测验的标准化工作中常常发生。如，某个企业引入一项新技术，所有员工都学习掌握该技术。经过一段时间培训后，所有员工掌握新技术的水平呈正态分布，因为有的人掌握得好，有的人掌握得差，大多数人水平中等。但在测查时，由于安排从不同的部门抽取一定的员工进行抽查，得出的分数分布可能由于抽样原因而不呈正态分布。这就需要校正。这一正态化过程主要是先将原始分数转化为百分等级，再将百分等级转化为正态分布上相应的离均值，并可以表示为任何平均数和标准差，计算步骤如下：

①对原始分数按序由小到大排列，计算各分数占总样本量的累积百分比；

②在正态曲线面积表中，求相对于该百分比的  $z$  分数。对大于中数的分数 ( $CP > 0.500$ ) 使用“较大部分的面积”一行的数据；对小于中数的分数 ( $CP < 0.500$ ) 用“较小部分的面积”一行的数据。所得的  $z$  分数可将分布分成几部分，我们称此值为  $z'$ ，以区分于由线性转换求得的  $Z$  分数。

③可以再次将分数转换成  $T$  分数量表，即平均数为 50，标准差为 10：

$$T = 50 + 10z' \quad (14.9)$$

④假如原始分布呈正态，正态化的标准分数与由线性转换所得的标准分数有相同的值。假如分布呈正态，这两种分数的值则不同。

以表 14—1 的数据为例，分数 77.5 的百分等级为 99.4，其累积百分比则为 0.994，查表，0.994 面积以下的  $z$  分数值是 2.51，若再将  $z$  转换为  $T$  分数，则：

$$T = 50 + 10z = 75$$

$T$  分数（T score）一词最早是由麦克尔 1939 年提出以纪念推孟和桑代克的，不过当时仅用于 12 岁的儿童团体，是根据某一特殊常模样本而不是在一般意义上定义的。现在有人用  $T$  分数来表示任何正态化与非正态化的转化标准系统，只要其平均数为 50，标准差为 10 即可。

（2）标准九（standard nine）分是另一较知名的标准分数系统，其量表是一个 9 级的分数量表。它以 5 为平均数，以 2 为标准差。标准九即标准化九分制的简称，它广泛用于美国空军和某些教学情境中的分级。

标准九将原始分数依据百分等级区分成 9 个等级，最高分为 9 分，最低分为 1 分，5 分位于分布的中心。除 1 分和 9 分外，其余每个分数均包括半个标准差的范围。

在一个正态分布中，每一个标准九所包含的百分比为：

标准九	1	2	3	4	5	6	7	8	9
百分比（%）	4	7	12	17	20	17	12	7	4

在使用标准九时，我们只要将百分等级分布上最高 4% 的被试给予 9 分，其次 7% 的被试给予 8 分，依次类推即可。当将原始分数转换成标准九时，事实上是将其置于一等距量表水平上，即 8 分与 7 分、5 分与 4 分以及 3 分与 2 分间的差距均相等（以标准差为单位）。这种标准量表使我们比较被试在不同类型测量上的相对地位的高低。

标准九的另外一个功能是可以给各种不同类型的资料予以加权，从而得到一个组合分数。例如，

先后有三种测验，我们认为最后一种测验的重要性是前两种的 2 倍，若已知第一次测验的标准九分为 6，第二次为 4，第三次为 8，则组合分数为：

$$\text{组合分数} = \frac{\text{加权分数总和}}{\text{权数总和}} = \frac{1 \times 6 + 1 \times 4 + 2 \times 8}{1 + 1 + 2} = 6.5$$

这个分数可以整体地代表被测者在三项测验上的总体水平。如果将所有被试的组合分数计算出来，就可依据整体成绩来给被试排定顺序。这也就是教学情境中分级的应用。然而，我们必须特别注意：组合分数并不是一种标准九分，故只能排定顺序。

其他常态化标准分数还有标准十分（卡特尔的 16 人格因素中使用）和标准二十分。前者平均数为 5，标准差为 1.5；后者平均数为 10，标准差为 3。

## 第十四章 人事测量结果的解释

### 第二节 效标与基于效标的解释

#### 一、内容参照分数

（一）内容参照分数（content referenced score）的测量目的是确定应试者对某个确定材料内容或技能的掌握和熟悉程度的分数。由于比较的对象不是其他人，而是掌握的内容多少，所以称为内容参照测验。成就测验就是典型的内容参照测验。在人事管理中，常常会用到这种类型的测验。例如，为了解新员工是否掌握培训班所授知识或岗前培训是否达到预定要求而进行的测验，就是内容参照测验。

（二）在编制内容参照测验时，要首先确定待考察的内容或技能，再编制确定应试者掌握程度的测验量表。编制内容参照测验量表的关键是预先制定一个判断应试者是否已掌握某种内容或技能的熟练程度的标准。这种标准可以通过掌握分数和正确百分数来表示。在实际中，对后者的应用更为常见。

##### 1. 掌握分数

最简单的掌握分数的标准是定一个判别应试者是否通过或掌握的最低分数。在此分数之上，表明应试者对考核的内容已经掌握，反之，说明应试者没有达到应该掌握的水平。通常以 80%—90%的人能

通过的分数作为最低分数。

2. 正确百分数

正确百分数表明被试在测验中答对题目的比例。计算公式如下：

正确百分数 =  $\frac{\text{答对题目数}}{\text{题目总数}} \times 100\%$  (14.10)

编制测验时，首先确定达到掌握或熟练标准的百分数，然后可以通过将应试者的答对百分数与此标准进行比较来确定该应试者的水平。

二、结果参照分数的解释

结果参照分数（outcome referenced score）是将效标材料直接结合到测验结果的解释过程而进行评价的分数。预测性测验往往适合用结果参照分数进行解释。结果参照分数可以表示获得某个分数的应试者达到某种效标水平的可能性。例如，在高级管理技能测量中，它的结果的解释可依据一定管理业绩或成效为标准。对于测验所得到的某个分数，可以判定它达到某一水平的绩效的可能性有多大。

进行结果参照分数解释的常用表示方法是期望表。期望表说明了一个给定的原始分数或分数等级获得不同效标分数或等级的可能性有多大，这种可能性用人数百分比来表示，如表 14—2 所示。

表 14—2 期望表示例

原始 分数	效 标 (%)							
	I	D		C		B		A
	0.0—1.9	2.0—2.9	3.0—3.9	4.0—4.9	5.0—5.9	6.0—6.9	7.0—7.9	8.0—8.9
75—79						100	100	67
70—74					100	100	72	35
65—69				100	94	30	33	11
60—64				100	85	48	22	4
55—59			100	88	63	31	4	
50—54		100	94	83	45	12	5	
45—49		100	87	61	24	5		
40—44	100	94	72	40	20	4		
35—39	100	85	46	15				
30—34	100							

从表中数据可以看出，该测验的效标分数为 10 分，共分为 A（优秀）、B（良好，又可分为  $B^+$  和  $B^-$ ）、C（合格）、D（较差）、E（很差）五个等级。表中数据表示与每一原始分数相对应的达到某一等级效标的人数百分比。例如，原始分数为 65—69 的应试者中，有 94% 获得了 C 等级中 5.0—5.9 这一效标分数  $C^+$  的水平，即获得 65—69 分原始分的被试有 94% 的可能能够达到该效标水平。

## 第十五章 人事测量在现实中的应用

### 第一节 人事测量工具组合的一般程序

测量工具的组合设计一般包括七个步骤：确定考评目的、需求分析、确定测量手段、预期结构、实施过程的设计、考评时间、费用预算或报价。

#### 一、确定考评目的

考评目的的确立，从大处讲要参考社会、市场、经济形势的发展状况，结合企业的长远发展战略、企业的文化追求。从小处讲，考评目的的确立要结合企业的经营策划、组织策划、变革策划。具体来说，要结合具体的人事管理目的和人力资源开发需要，考虑内部的具体需求和特殊动机等多方面的信息，来确定人才素质测评的目标和方向。

#### 二、需求分析

需求分析建立在对施行人事测量的企业有全面、深入的认识基础上。这是一个分析与归纳的过程，需要人事测量专业人士与企业双方面的广泛深入沟通来达成共识——甄别人员素质考评要素。人员素质考评要素必须反映企业文化和理念的要求、岗位工作的要求、岗位职责的要求、个人可发展性的要求、适应性要求。

从内容上看，考评要素包括个性品质、职业适应性（如兴趣取向、价值取向、需求与动机水平）、技能与能力。其中，技能与能力包括：资格认证水平（如专业技能水平、外语、计算机应用水平等）、基本能力（如智力水平）、职业能力（如营销技能）和管理技能（如计划、预测、决策、沟通能力）。

在具体操作中，首先要在细致全面的工作分析之上，具体确定每一职位所要求的能力、个性品质



（性向、情绪、风格）以及职业适应性和匹配度这三方面的要点，并且构建各个要点之间的整合关系，据此作为考评要素，它也是下一步组织测验方案的评价目标，同时也是做评价结果报告时各项测验得分进行综合评估和优劣评定的参照依据。

### 三、确定测量手段

有了详细的需求分析，就可以选择具体的测验工具项目了。在每选择一个具体的工具时，要说明这个测量工具的功能，采用的理由。在选择测量工具时，必须考虑到企业的特色与文化。需要注意的是，对于难于控制掩饰倾向的考评要素，可能的话，应选择使用多项测验工具，这样可多次考察该要素，在结果分析中可多方参照，相互印证，了解该方面的真实情况。

### 四、预期结果

确定测量工具后，应重新系统地评价这个设计方案，对可能的结果作出预期。作预期时应该考虑以下三个方面的内容：

#### （一）由具体的分数如何建立综合报告

由于有许多要素要考察，同一要素又可能由多种方式（工具）考察，因此会得到大量测量分数，如何将这些数据整合在一起，如何根据测量分数撰写综合报告？不同分数之间是否会出现矛盾以及如何处理这些矛盾数据？这些在设计时都应有所考虑。

#### （二）测量的结果将如何指导后期的工作

测量的结果会因测量目的的不同而对工作具有不同的意义。针对招聘的测量会影响到人力资源的质量；针对培训的测量会影响培训决策和效益；针对晋升的测量会触及个人和企业双方的利益。此外，由于目的不同，测量标准可能服务于择优或淘劣，在具体标准的设定上也有所不同。

#### （三）可能在更大范围中对员工产生的影响

最后，还要考虑测量的实施会在更大范围中对员工产生的影响。测量方案是否对所有人都公平？是否适合各个阶层的人？人们会接受它吗？还是会出现很大抵触？时间上是否安排得开？是否会产生疲劳和厌倦？是否会影响正常工作？投入（时间、精力、资金等）—产出比如何？员工会如何反应？

是否需要事先舆论和技术上的准备？所有这些问题都应当有周密的考虑，并在设计方案中给予说明。

## 五、实施过程的设计及原则

### （一）实施的一般过程

现在要把实施所有测量的具体程序确定下来，包括说明有关时间、地点、辅助材料、现场布置要求、分组（如果需要的话）、设备、流程等所有细节工作的安排、落实。

### （二）设计的原则

1. 简便易行的测量放在前面。
2. 成本低的测量放在前面。
3. 当一个测量的内容可能影响（如暗示、帮助）其他测量时，这个测量应放在后面。
4. 容易产生疲劳的测量放在后面。
5. 测量内容比较敏感，或容易造成较大压力的测验应放在后面。

## 六、考评时间的确定

测量组合设计时要考虑和说明测量项目的时间。有时，由于特殊需要，测量项目会比较多，总时间就会很长，有时甚至会需要几天的时间。这时应当有成本—收益分析，在过长测评的疲劳、对正常工作的影响与科学、系统的诊断之间找到一个恰好的平衡点。许多组织常常是因为对系统的测量的时间不能接受而放弃测量计划的。因此，可行性也是需要考虑的因素。

## 七、费用预算或报价

在完成一个人事测量的设计时，要给出它的预算（对组织内部而言）或报价（就组织外部而言）。

在人事测量行业，测量的费用通常按照每个人每项测量内容来计费的。另外，当需要对一个人的多项测量结果给出综合报告时，会有额外的费用。这通常会发生在针对管理人员或专业技术人才的测量费用里。

在西方，同时测量大量人员往往会比同等人数单人测量的合计费用要高。这是因为专业人员一下子要处理、分析大量测量结果会有相当困难，也会有疲劳，甚至影响其分析质量。

当一个组织同时测量很多人，特别是整个组织大规模测评时，往往会需要对所有被测量人员的综合结果，或是对整个组织的情况作出全面报告。这个报告总是会被额外计费的，因为它要求额外的系统分析。当然，这个工作是很有价值的，它给出了组织整体人力资源的概貌。

## 第十五章 人事测量在现实中的应用

### 第二节 针对不同岗位系列的人事测量

#### 一、企业的基本运作单位

一般来说，企业组织的基本运作单位包括生产（创造可使用的东西或增加商品和服务效用）、营销（寻找和吸引愿意按价格购买商品和接受服务的顾客或建立分销渠道）、财务（收支、保管和运筹企业资金），以及行政人事（安排人力资源并完成各部门之间以及内外部的必要协调）等。

#### 二、各岗位系列所要求的基本素质

##### （一）生产岗位系列基本素质要求

1. 能力：时间管理（有时间概念，讲效率，严格保证生产进度）、有效的控制力、系统全面的计划能力；

2. 个性：认真、负责、准确；

3. 知识经验和技能：专业水平高，熟悉生产流程和相关设施，具备安全生产知识。

##### （二）营销岗位系列基本素质要求

1. 能力：良好的沟通、协商、应变能力和人际能力，适应力强，富有创造性，情绪控制能力；

2. 个性：乐群、热情、耐心、坚韧执着，能承受挫折感，健谈；

3. 知识经验和技能：最好有客户关系，熟悉分销渠道，具备营销常识。

##### （三）财务岗位系列基本素质要求

1. 能力：判断力、决策力、金融预测能力；

2. 个性：细心、严谨、准确，为人谨慎、有序、负责、原则性强；

3. 知识经验和技能：有专业资格认证、经验越丰富越好。

#### （四）行政人事岗位系列基本素质要求

1. 能力：人际沟通能力、适应能力、全面细致的分析能力；
2. 个性：随和、热情、友善、令人信赖，细致耐心；
3. 知识经验和技能：熟悉有关劳动法规，具有相关工作经验。

#### （五）技术岗位系列基本素质要求

1. 能力：独创性，发现和解决问题的能力，思维严密善于学习；
2. 个性：自信、创造性、追求自我实现，专注于技术进步，最好具有研究价值取向；
3. 知识经验和技能：技术等级认证，相应技术领域工作经验，研究开发的理论技术成果。

### 三、各岗位系列测量工具组合

表 15-1 列出了各岗位系列主管人员的主要考察要素和供选用的测量工具。至于各岗位系列的普通员工，可参照下表，减去和管理工作有关的内容，或替换某些内容，如把管理个性测验换成适用于普通人的 16PF 人格测验。

**表 15—1 不同岗位任职要素的测量工具**

岗 位	考察要素	考评核心工具
生产系列	个性特征、组织协调能力、综合分析能力、兴趣取向、行为风格、工作经历	管理个性测验或 DISC 测验 兴趣偏好测验 价值观评定 面试（结构化或半结构化）
营销系列	人际敏感性、沟通能力、个性特征、动机需求模式、言语表达、工作经历	管理个性测验或 DISC 测验 敏感性与沟通能力测验 需求测试 生活特性问卷 无领导小组讨论 面试（结构化或半结构化）
财务系列	个性特征、思维分析能力、综合决策力、工作经历	管理个性测验或 DISC 测验 数量分析能力测验 面试（结构化或半结构化）
行政人事系列	个性特征、人际技巧、事务处理能力、工作经历	管理个性测验或 DISC 测验 无领导小组讨论 领导行为评定 面试
技术系列	创造性、逻辑推理能力、个性特征、工作经历	管理个性测验或 DISC 测验 逻辑推理测验 抽象推理测验 面试（结构化或半结构化）

## 第十五章 人事测量在现实中的应用

### 第三节 针对不同企业特征的人事测量

#### 一、针对企业行业特征的组合分析

企业所从事的行业不同，对其从业人员的素质要求也不尽相同。由于行业性质的不同，各自形成自己的行业特点和规范，这些特点和规范反映了对从业人员的能力、个性特征、适宜的动机需求结构等各个方面的要求。这些要求的侧重点不同，因而对不同行业人员的测评内容和考察因素都需要进行具体分析，在测验设计上应反映出行业的特点。

表 15—2 列出了对一些具有突出行业特征的人员特殊性需求分析，并列出在人事测量中针对这些

行业特征要求，可以实现考察目的的考评工具类型。

**表 15-2 不同行业特殊需求及相应测量工具**

行 业	需求分析	测量类型
生产制造业	全面严格的质量控制力 创新开发能力	个性测验
服务产业  文化产业	善于服务取向的个性、关键 人际技能  创造性思维 高强度组织规划能力 综合能力	个性测验 人际技能测验  思维测验 管理能力测验 案例分析
高新技术产业	独立性 学习能力 科技敏感力 高新技术造诣 敏锐与信息把握和应变能力	思维测验 情境模拟测验 案例分析

二、针对企业规模特征的组合分析

企业规模的大小直接反映在企业的组织结构和人员组成上。

规模小的企业组织结构简单，层次少，组织成员少，往往人事管理职能不完善。对这类企业的应用测评可在基础性的人事职能上做一些替代性的工作，如通过测评来决定人员录用和工作分配、晋升等，因此在测验设计上应面向个人做比较详细的测查，必要时可做包括面试在内的一整套人员甄选程序。对小规模企业一般不做团体层次或组织层次的测评。

规模大的企业，组织结构复杂，人员数量大，人事管理系统比较健全，往往需要通过测评来收集一些辅助性的信息，比如调查员工的工作满意度、员工动机需求状况，为制定有效激励政策提供参考；还可以测查团队绩效和团队健康度，为实现高效的团队合作和提高团队成长力服务。当然，并非所有的大规模企业都有完善的人事管理部门，能够执行全面的人事管理职能。对于这样的大型企业，对人员测评的需求同样是巨大的，测评服务的管理咨询功能将大有可为。

三、针对企业文化特征的组合分析

任何企业都有自己区别于其他企业的独特性文化。企业文化是企业中长期形成的共同思想、作风、

价值观念和行为规范，是一种具有企业个性的信念和行为风格。

不同企业在企业文化上的重视、培育力度不同，其企业文化影响程度就不同。在对不同企业进行测验设计时，有必要了解其经营理念和追求，了解企业文化的建设状况，据此有针对性地使用不同内容的测验，使测评的结果更有实用性和应用价值。

# 全国职业经理MBA双证班

**认证系列：**职业经理、人力资源总监、营销经理、品质经理、生产经理、物流经理、项目经理、企业培训师、酒店经理、市场总监、财务总监、行政总监、采购经理、营销策划师、企业管理咨询师、企业总经理等高级资格认证。

**颁发双证：**高级经理资格证书+MBA 高等教育研修结业证书（含 2 年全套学籍档案）

**证书说明：**证书全国通用、电子注册，是提干、求职、晋级、移民的有效依据

学习期限：3 个月（允许工作经验丰富学员提前毕业） 收费标准：全部学费 **1280** 元

**咨询电话：** 13684609885    0451- 88342620    **招生网站：** <http://www.mhjy.net>

**电子邮箱：** [xchy007@163.com](mailto:xchy007@163.com)    **颁证单位：** 中国经济管理大学    **承办单位：** 美华管理人才学校

## 全国招生 函授教育 颁发双证 权威有效



美华论坛  
[www.mhjy.net](http://www.mhjy.net)



职业经理 MBA 整套实战教程

千本好书 **免费** 下载    学校网址： [www.mhjy.net](http://www.mhjy.net)