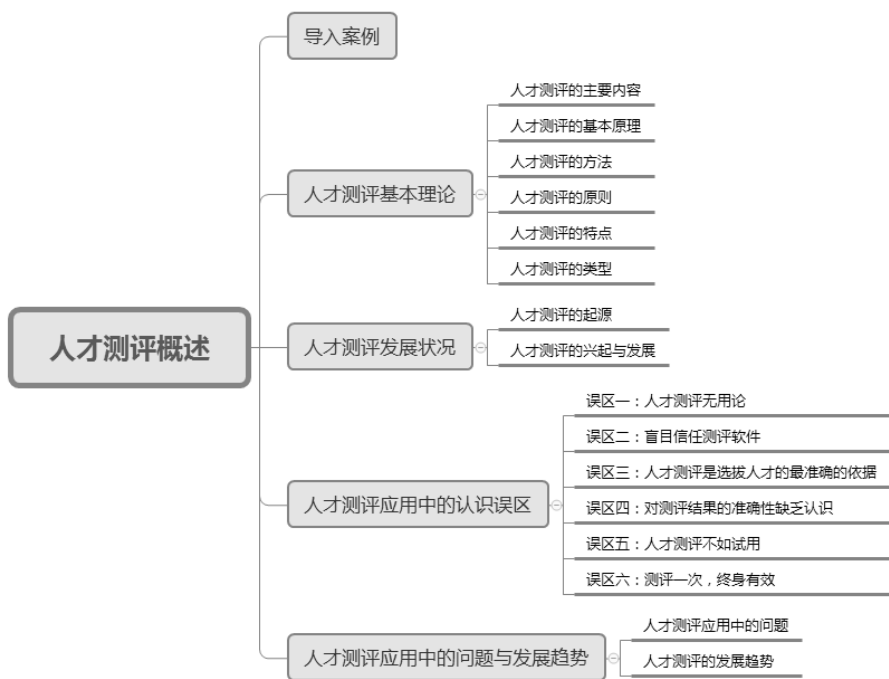


# 第1章

# 人才测评概述

本章思维导图



## 本章关键词

人才测评；基本理论；发展状况、认识误区、发展趋势。

### 学习目标

- 了解：人才测评的概念，人才测评的发展状况、兴起与发展。
- 熟悉：人才测评应用中的认识误区。
- 掌握：人才测评的发展趋势。

### 导入案例

#### 某通信公司用人才测评进行精准的培训需求分析

深圳某通信公司，主营业务是网络通信终端设备的研发、生产、销售与服务。该公司在经历了高速发展期之后，开始进入稳定发展阶段，公司的中层管理人员绝大多数是由业务核心骨干提拔上来的。随着公司管理体系的日益完善，这些非管理出身的管理者在管理岗位上越来越觉得吃力。虽然公司也在培训中引入了一些流行的领导力课程，但是实际效果非常有限。因此，公司决定对中层管理人员进行更为精确和更有针对性的培训需求分析。

##### 一、顾问观点

传统的企业培训需求分析，通常运用访谈法、问卷调查法、绩效分析法、阅读技术手册法、访问专家法、经验预测法、头脑风暴法进行，只关注组织、任务、个体三个基本要素，导致企业培训工作在提高员工个体及企业整体优势作用方面表现不理想。

基于胜任特征和人才测评的培训需求分析，通过分析组织核心能力、构建企业岗位胜任特征模型、评估现有任职人员胜任能力三大步骤，测评个体与胜任特征模型中行为要求之间的具体差距。据此发现被试者（培训对象）整体的能力“短板”，从而把对被试者的培训重点放在其具体的行为改进和技能提升方面。

##### 二、实施方案

###### （一）确立项目目标

确立项目目标是指对该公司中层管理人员实施培训需求分析诊断，为中层管理人员后期的培训重点提供科学参考。

###### （二）项目准备

###### 1. 成立工作小组

由公司人事部牵头，邀请测评公司专业顾问一起成立分工明确的工作小组。

###### 2. 确定培训和测评对象

公司培训和测评对象确定为中层管理人员，共28人。

### 3. 确定组织需求和个人需求两大维度

在组织需求方面，从组织发展的角度，确定公司中层管理人员需要哪些能力，目前哪些方面的能力需要提升；在个人需求方面，从员工个人职业成长规划的角度，确定需要提升哪些能力。

#### （三）构建中层管理人员胜任特征模型

工作小组在测评顾问的指导下运用行为事件访谈法、问卷调查法、专家讨论法，使用 SPSS 和 AMOS 数据分析软件，并根据公司中层管理人员的岗位职责、企业文化的特殊要求、企业发展规划和发展战略等，对企业中层管理人员的素质要求等方面，确定胜任特征模型的指标、维度和内涵。据此，构建了通信公司中层管理人员胜任特征模型。

#### （四）现有任职人员的胜任能力测评

##### 1. 确定标准分数

基于企业战略、企业目标、企业价值文化、优秀员工的素质的交集，整合确定中层管理人员的标准分数，作为人才测评的参照系。

##### 2. 确定测评方法

胜任能力测评方法采取在线情景判断测验，即给被试者呈现他在工作中可能遇到的各种情景，这些情景通常来自关键性事件和工作分析，伴随着每种情景的，是对假设情景的各种可能的处理方法或方式，然后要求被试者做出判断（选择最可能/最不可能的），对照最佳答案对其胜任能力进行评估。

情景判断测验的特点是：①题目场景来源于工作实际，真实反映胜任能力，可以较好地预测绩效；②每个选项都是解决情景问题的一个可能的办法，只是重要性和有效性有差别，不易看出好坏，能够刺激被试者深入思考，做出最符合内心判断的选择；③封闭式题目，预设评分标准，保证测评公平公正，不受主观因素影响；④借助测评公司提供的成熟的测评平台，使用简便、易于操作，结果反馈较快。

##### 3. 测评题目开发

工作小组根据测评要素，收集岗位工作任务与员工关键行为方面的信息，参考测评公司提供的测评题库，进行相应的题目开发，也就是命题工作，同时设计评分表及评分标准。在此过程中，要进行测评题目难度、区分度、信度、效度等的检验或评估。

##### 4. 测评流程设计

根据参与测评的中层管理人员的数量，安排测评时间，设计测评流程，协调相关资源。

### 5. 前期宣导

为了保证测评工作进行顺利，由组织者对被试者进行宣导，主要告诉他们正确看待本次测评，认真答题，不能故意掩饰或随意作答。

### 6. 测评过程实施

由测评组织实施人员在测评系统上生成测评任务二维码，让被试者用手机扫描二维码，填写基本信息，开始作答。作答完成后，系统自动回收数据，组织者可以即时查看每个人的测评报告和得分汇总数据。

#### (五) 测评成果及应用：构建“学习地图”，制订培训计划

根据系统自动生成的测评结果，得到每个人的测评报告。

根据系统提供的 Excel 版得分汇总数据，做出组织的整体报告。

通过科学客观的评估工具，使公司领导者认识到，中层管理人员的普遍“短板”是在沟通协调能力和团队领导能力相关指标上，通过制订针对性的培训方案，对中层管理人员进行培训，以提高其管理能力和素质，并为人力资源的培训开发提供测评依据。

通过本次中层管理人员培训需求测评项目分析，能够对后备干部的能力素质进行摸底，择优选拔具有高潜力的中层管理人员，为其制订具体的职业发展规划和培养方案。同时，根据中层管理人员个人测评成绩，可针对其个人短板，制订个性化培训计划。

根据胜任特征模型，构建中层管理人员“学习地图”。以员工能力素质要求为基点，规划培训体系。让公司领导者清楚中层管理人员的整体水平状况；让中层管理人员增强自我认知，发现和寻找自身的不足，扬长避短，制订个人的学习计划。

(资料来源：<https://www.beisen.com/customer/#top-list-end>)

---

## 1.1 人才测评基本理论

---

### 1.1.1 人才测评的主要内容

素质主要包括品德、心智、能力、文化、身体素质等方面，其主要内容如图 1-1 所示。

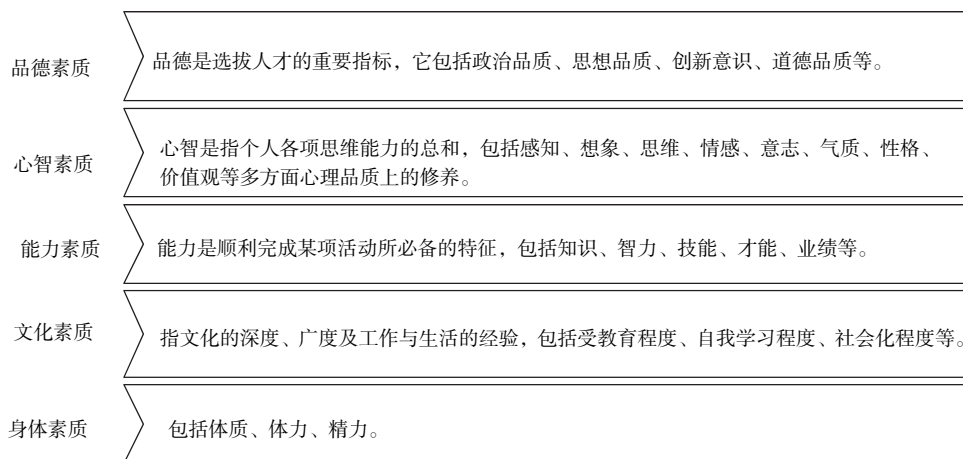


图 1-1 素质的构成要素

人才测评主要考查具有稳定性的个人素质，根据素质的构成要素，其考查内容包括个人的品德因素、能力因素、动力因素和人格因素。

### 1. 品德因素

品德因素包括政治品质、思想品质和道德品质等，它是个人在社会化过程中形成的稳定的心理特征和倾向，是个人遵守的社会规范。品德水平的高低直接影响了个人在社会情景中的价值选择。

### 2. 能力因素

能力因素包括知识、智力、技能、才能、业绩、创新意识等，可以将其划分为科学智能和社会智能。其中，科学智能包括人与自然交往中获得的经验或个人通过书本学习到的间接经验；社会智能则来自个人在社会实践中，通过人与人之间的沟通、联系、竞争与合作获得的经验。

### 3. 动力因素

动力因素包括价值观、动机和兴趣等。

(1) 价值观是指一个人对周围客观事物（包括人、事、物）的意义、重要性的总评价和总看法，它是层次最高、影响力最大的因素。

(2) 动机是指由特定需要引起的，欲满足各种需要的特殊心理状态和意愿，它是推动一个人行为的内在原因，动机的强烈程度对行为过程的效率和行为的结果有很大的影响。

(3) 兴趣是指个体对某种活动或某种职业的喜好，当个人的兴趣与行动方向一致

时,可以使行为更加有效。

#### 4. 人格因素

人格因素是个体在行为上的内部表现,是个体在适应环境中的感情、能力、气质、价值观等方面的内部综合表现,它可以离开人的肉体、离开人所处的物质生活条件而独立存在于人类的精神文化维度里。

人的个性差异表现为四种气质类型,包括多血质、胆汁质、黏液质和抑郁质。另外,人格测验可以从四个方面考查人的行为风格:一是一般心理倾向(外倾和内倾);二是接收信息的方式(感觉和直觉);三是处理信息的方式(思考和情感);四是行为方式(判断和知觉)。

### 1.1.2 人才测评的基本原理

#### 1. 人才测评的理论基础

现代人才测评技术出现于 20 世纪初,并且作为一门专业化学科随着现代经济发展、科技进步而迅速发展起来,众多相关学科诸如心理学、心理测量与咨询、统计学、管理学,以及计算机技术的发展,都为人才测评得以实施奠定了可行性理论基础。

(1) 素质差异性。众所周知,人与人之间存在生理上的差异,性别、外貌差异是显而易见的,但是根本性的心理差异是难以衡量的。心理学领域的专家通过长期的研究,把个体之间的差异总结为相互关联的两个方面:一是个体的倾向差异,包括兴趣、爱好、需要、动机、信念、理想、世界观等方面的差异;二是个性心理特征差异,包括能力、气质与性格三大因素及其组合的差异。正是这种个体差异,才使得人才测评得以存在且意义重大。因为人才测评能够将这种千差万别的个体差异准确反映出来,为人力资源开发提供决策指导。

(2) 素质稳定性。素质是一种相对稳定的组织系统,每个个体不尽相同,表现为人在不同环境下、不同时段内对某些不同的刺激做出一致的反应行为。因此,也可以说,素质的稳定性体现为一个人某种经常的和一贯性的行为特点。比如,某人对某事在各个时间段的反应基本一致,就表现为跨情景的一致性。这二者就构成了素质的稳定性。但是,素质的稳定性具有相对性,个体的素质会随着时间推移而发生变化,而这种变化时间往往很长。正是这种稳定性,人们才能掌握个体的素质,人才测评才能够得以实施。

(3) 心理测量学。心理测量学的产生和发展对人才测评技术有着至关重要的作用,量化使人才测评得以表现测评各项目的数量特征和质量特征,从而避免主观性的偏差。从 1879 年德国心理学家冯特设立世界上第一个心理学实验室开始,人们便对个体行为

差异进行研究,心理测验运动也随之兴起。心理测验由早期的心理缺陷诊断演变为心理评价,由教育领域逐渐拓展到社会管理的各个角落,给升学、就业、选拔人才、晋升、培训等提供指导性服务,得到西方社会的普遍欢迎与重视。20世纪五六十年代以后,西方人才测评思想盛行,开发了一系列形式多样、内容丰富的测评技术,包括智力测验、能力测验、性向测验、成就测验等,并且这些技术在广泛应用的基础上日趋完善。目前,人才测评中最常见的心理测验有智力测验、人格测验、成就测验和职业倾向测验。另外,随着统计学与计算机技术的发展,专门化的测评软件使测评的客观性与效率大幅度提高。

## 2. 人才测评应用的理论基础

作为一门应用型学科,人才测评已发展出一套专门的理论与方法,以此指导人才测评实践的操作。

(1)“人职匹配”理论。人职匹配理论是关于人的个性特征与职业性质是否一致的理论,也可以说是人才测评与工作分析相结合的理论。其基本思想是,个体差异是普遍存在的,每个个体都有自己的个性特征,而每种职业由于其工作性质、环境、条件、方式的不同,对工作者的能力、知识技能、性格、气质、心理素质等有不同的要求。因此,在进行职业决策(选拔、安置、职业指导)时,就要根据一个人的个性特征来选择与之相对应的职业种类,即进行人职匹配。人职匹配包含两层意思:一是指某个人的能力完全能胜任该岗位的要求,即人得其职;二是指某个人完全具备岗位所要求的能力,即职得其人。人职匹配理论指人的能力与岗位要求的能力完全匹配,这种匹配包含着“恰好”的概念,二者的对应使人的能力发挥得最好,岗位的工作任务也完成得最好。人职匹配理论的核心因素是:最优的不一定是最匹配的,最匹配的才是最优的选择。即职得其才,才得其职,人职匹配,效果最优。所以,对于组织和个体来说,进行人职匹配具有非常重要的意义。进行人职匹配的前提之一是必须对人的个体特性有充分的了解和掌握,而人才测评是了解个体特征最有效的方法。可见,人职匹配理论是现代人才测评的理论基础,其中最具有影响力的是“特性—因素论”和“人格类型理论”。

特性—因素论认为个别差异现象普遍存在于个人心理与行为中,每个人都具有自己独特的能力模式和人格特质,而某种能力模式及人格特质又与某些特定职业有关。每种人格特质的个人都有其相适应的职业,人人都有选择职业的机会,人的特性又是可以客观测量的。特性—因素论强调个人所具有的特性与职业所需要的素质与技能(因素)之间的协调和匹配。为了对个体的特性进行深入详细地了解与掌握,特性—因素论十分重视人才测评的作用。可以说,特性—因素论进行职业指导是以对人的特性的测评为基本前提的,它首先提出了在职业决策中进行人职匹配的思想。

美国职业心理学家霍兰德创立的人格类型理论对人才测评的发展产生了重要的影响。在人格和职业的关系方面，霍兰德提出了一系列假设：①在现实的文化中，可以将人的人格分为六种类型：实际型、研究型、艺术型、社会型、企业型及传统型。每一特定类型人格的人，会对相应职业类型中的工作或学习感兴趣；②环境也可区分为上述六种类型；③人们寻求能充分施展其能力的职业环境；④个人的行为取决于个体的人格和所处的环境特征之间的相互作用。在上述理论假设的基础上，霍兰德提出了人格类型与职业类型模式。不同类型人格的人需要不同的生活或工作环境，例如，“实际型”的人需要实际型的环境或职业，因为这种环境或职业才能给予其所需要的机会与奖励，这种情况称为“和谐”。

据此，人力资源管理者便可根据人才测评结果，并结合职位要求，进行人事配置，使事适其人、人事相配，以求最大限度地开发人力资源。

(2)“冰山模型”理论。在人才测评过程中，人们一般比较重视考查知识、技能，以求与职位要求有效匹配。其实不然，成功的关键还要取决于被试者深层次的因素——价值观、态度、角色定位、品质和社会动机等，即“冰山模型”水平面下的关键要素。1973年，美国心理学家戴维·麦克利兰提出了著名的冰山模型理论，他把人的胜任特征模型形象地描绘成一座在水中漂浮的冰山(见图1-2)，“水上部分”代表表层的特征，如知识、技能，容易被感知；“水下部分”是潜在的特征，如价值观、品质和社会动机等，是决定人们行为和表现的比较稳定的关键因素。越向下，被挖掘与感知的难度越大。并且，一个人的表层特征只是整体特征的极少部分。因此，我们在进行人才测评时，应该从胜任特征入手，重点考查被试者的价值观、品质和社会动机等。只有这样，才能把真正高素质的人挑选出来。

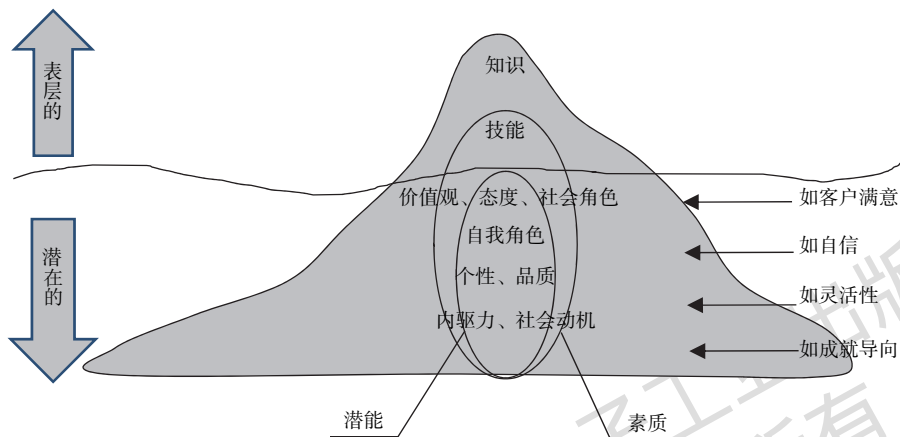


图 1-2 戴维·麦克利兰的“冰山模型”



### 1.1.3 人才测评的方法

俗话说：“路遥知马力，日久见人心。”但是“马力”易知，“人心”难测。所以，要想实施人才测评，就必须先掌握测评方法，正所谓“工欲善其事，必先利其器”。人的心理是复杂多变的，还会受到各种外界因素的影响，加上人们有意地掩饰，要想将人才特质准确、客观地测量出来绝非易事，甚至本人对自身都难以做出客观的评估。所以，必须采用科学有效的方法才能做到“测人心”。常用的人才测评方法主要有以下三类。

#### 1. 心理测验

心理测验是人员素质测评与选拔的最常见的形式之一，实质上是对行为样组的客观的和标准化的测量。心理测验产生于对个别差异鉴别的需要，广泛应用于教育、企事业单位的挑选与评价，具有迅速、科学、公平和可比较的特点。从测验的具体对象出发，可以将心理测验分为成就测验、智力测验、性向测验、态度测验、性格测验和品性测验六种形式。依据测验的内容，可以分为知识测验、能力测验、品德测验三种形式。其中比较有影响力的心理测验，除了早期的比奈-西蒙智力测验、斯坦福-比奈儿童智力测验，还有罗夏墨迹测验、亨利·默里与摩根的主题统觉测验（Thematic Apperception Test, TAT）、明尼苏达多相个性测验（Minnesota Multiphasic Personality Inventory, MMPI）、艾森克人格问卷（Eysenck Personality Questionnaire, EPQ）、卡特尔 16 种人格因素测验（Catell 16 Personality Factor Test, 16PF）、皮亚杰故事测验、科尔伯格两难故事测验、雷斯特确定问题测验等。这些测验所采用的技术，都要尽可能排除主观因素的影响，如机器评分，采用简答、填空、选择等客观性试题，但是由于心理的复杂性，心理测验也不能完全客观化。

纸笔测评是最原始也是最常见的心理测验形式，心理测验中的人格测验、能力测验、兴趣测验，一般都采用纸笔测评的方式，要求被试者填写答题卡，或在卷面上直接完成题目。由于纸笔测评在人才测评中的重要作用，大部分测验都要求被试者直接在卷面上写出答案，考查人才的智力、知识、能力及发展倾向。从 1988 年开始，我国中央国家行政机关公务员考试行政能力测试科目就一直采用笔试形式。纸笔测评的优点在于科学化、规范化的水平较高，适合大规模的人才测评，测试结果也比较令人信服。但是，笔试内容只局限于文字作答，在检测被试者解决复杂的实际问题的能力方面却鞭长莫及。

#### 2. 评价中心

评价中心是综合多种人才测评方法，以定性定量相结合的方式对被试者进行测评的一套人才测评程序。其特点是综合性、动态性、标准化、整体互动性、全面性、以预

测为主要目的、形象逼真,以及具有很强的行为性。由于其具有较高的预测效度,目前评价中心在应用方面呈现增长的趋势。美国匹茨堡大学职业研究院的威廉·C·帕海姆教授在调查了评价中心的研究项目后指出,经过评价中心测试选拔的管理人员比仅凭主观判断而提拔的管理人员,其成功率要提高2~3倍。20世纪50年代,美国电话电报公司最早成功使用评价中心,获得了公司中、高级经理及候选人的特点和成长状况方面的资料。1975年,在加拿大魁北克召开的第三届评价中心国际大会上通过了《关于评价中心的实施标准和道德准则》,随后各国相继展开了评价中心本土化的研究。20世纪90年代以后,伴随中国经济的腾飞和改革开放的深入,评价中心逐步应用于企业人力资源管理工作中,在公务员的招聘录用中,评价中心也发挥着越来越重要的作用。评价中心的形式包括公文筐测验、无领导小组讨论、管理游戏、角色扮演、演讲、案例分析、事实判断、面谈等,我国评价中心各种形式使用频率,如表1-1所示。

表 1-1 我国评价中心各种形式使用频率

复杂程度	评价中心的形式	实际运用频率
复杂 ↓ 简单	管理游戏	15%
	公文筐测验	89%
	角色扮演	64%
	无领导小组讨论	75%
	演讲	21%
	案例分析	83%
	事实判断	43%
	面谈	93%

资料来源:彭志忠,王永莲.人才测评学[M].济南:山东大学出版社,2006.

### 3. 计算机测评

随着计算机应用的普及和网络技术的飞速发展,人才测评的方式也焕然一新,与现代信息技术、网络技术紧密地结合到了一起。人们开发出了单机版、网络版的心理测验,这些测验不但包括诸如16PF的心理测验,还包括像“销售经理选拔测验系统”等针对特定岗位的测评系统。近年来,有些测评机构将互联网不受时空限制的特点发挥得淋漓尽致,将心理测验放到互联网上,用户可以随时开展测验工作,进一步拓宽了测评的渠道。人才测评需要在初步得出的基本参数上进行精细分析,才能得出一个人的个性特征的总体轮廓与具体评价,但电脑在人才测评中只能充当“评分者”的角色,完成统计基本参数的机械操作,纸笔测评比较适合“搬”到计算机上进行测评,而情景模拟测试很

难实现计算机测评。

#### 1.1.4 人才测评的原则

人才测评过程,必须遵循一些重要而基本的原则,这些原则既是人才测评实践经验及其技能技巧的科学总结,又是人才测评实践的思想方法。

##### 1. 定量测评与定性测评相结合

人才测评过程中主要存在两种研究方法:一是运用数学工具从行为的数量特点上对人才进行量化的测评,研究的目的在于确定因果关系,并做出解释;二是采取经验判断与观察的方法从行为的性质上对人才进行测评,注重的是整体和定性的信息。定量测评是定性测评的基础,定性测评是定量测评的出发点与结果。仅仅运用定量测评,人才测评的结果只能是一堆没有意义的数字和数量关系,成为一种不完整、形式化的测评;而仅仅使用定性测评,人才测评又会流于表面,成为一种简单、模糊的印象判断。所以,要把定量测评与定性测评结合起来,其实质就是将人力资源管理的经验与科学测量的方法相结合。

##### 2. 客观测评与主观测评相结合

客观测评是指在人才测评过程中尽量采用客观的工具进行测量、评定,减少主观因素的干扰;主观测评是指凭借专业知识和经验,通过对人的观察,做出主观的评定。在人才测评过程中,客观测评以客观的行为和事件为依据,尽量减少或排除测评人员和被试者的主观随意性,并保证不同的测评人员在对相同的被试者进行测评时,能做出大致相同的结论。但是,人的行为是复杂多样的,不是所有的素质都能靠客观测评得以把握。因此,在人才测评过程中追求客观化的同时,也应该充分发挥人的主观性,从人的主观方面对人才素质进行考量,即客观测评和主观测评相结合。

##### 3. 静态测评与动态测评相结合

在人才测评过程中,还要遵循静态测评与动态测评相结合的原则。静态测评是建立在人才素质稳定性的基础之上,在特定时空条件下以相对统一的测评方式对人才进行测评,人才要素的动态变化性忽略不计。静态测评便于对被试者实施横向比较,可用于选拔和储备人员,缺点是容易孤立看待被试者,难以预计今后的发展趋势。动态测评是从发展的角度看待人的素质,有利于了解被试者的实际水平与发展潜力,可用于激励和开发人力资源,但不利于横向比较。所以,如果想深入准确地把握被试者的素质就要静态测评与动态测评相结合,有的标准是以往大量实验研究得出的常模,具有标杆意义,可

以推广到各种人群,有的标准则要随着时代的进步不断更新。静态测评方法包括心理测验与知识考核;动态测评方法有评价中心、面试与观察评定等。静态测评简便易行,但易受人为因素干扰;动态测评有效性高,但人力、物力、财力和时间成本非常高,并且需要一定的技术条件和理论基础作为支撑。

### 1.1.5 人才测评的特点

人才测评是一项复杂的社会认知活动,具体的测评对象不是抽象的人,而是作为个体而存在的人的内在素质及其表现出的绩效,要受现实生活的种种制约,这就决定了人才测评不同于其他形式的测评活动。其特点主要表现在以下几个方面。

#### 1. 相对性

人才测评主要是对个体心理现象的测量和评价,包括智力、能力、性格、气质等方面。而心理现象具有内在性、隐蔽性和无形性等特点,再加上任何测评工具都难免有一定的局限性,对被试者的鉴别评价就未必完全符合被试者的实际情况,测评结果既反映被试者的基本状态,又与被试者的真实情况有一定程度的偏离。所以,人才测评的结果只有相对意义,在测评实践中应注重测评的精确性和科学性,不应片面追求精确性。

#### 2. 统计抽样性

从理论上讲,在实施人才测评时,涉猎的范围越广,收集的信息越充分、全面,测评结果就越有效、越具体客观。但在实际操作中,这种理想状态是不可能达到的。在有限的时间内不可能掌握被试者的全部信息,只能依据“部分能够反映总体”的原理,对测评要素进行抽样,保证样本有足够的代表性。因此,在测量个体的特性前,必须慎重地选择具有代表性的行为样组,凭此推论个体的特性。

#### 3. 互动性

人才测评不是简单的单向刺激反应过程,而是一种反复多次的双向互动与沟通的过程,是一种集理智、情感、行为于一体的互动过程。被试者既可以积极配合测评,也可以干扰、拒绝测评。所以,如果忽视互动性,没有调动被试者的积极性与配合性,即使测评内容与程序设计得十分科学,也不一定收到预期效果。因此,在人才测评中,尤其是面试环节,不能把被试者当作物品来测,而应力求让被试者由客体状态向主体状态转化,发挥其积极的作用。

## 4. 整体性

个人素质、能力及其绩效的好坏,从根本上说,都是与他人比较而得出的。对单个人进行“孤立性”的测评,往往难下结论,而把被试者放到整体被试者群中加以比较,其优劣高低就立见分晓了。因此,人才测评的内容和标准,要因被试者群的不同情况而有所侧重。

### 1.1.6 人才测评的类型

人才测评的类型,按不同的标准有不同的划分。

按测评标准划分,有无目标测评、常模参照性测评及效标参照性测评。述职、小结与访谈等写实性测评,属于无目标测评;人才录用与招聘、晋升一般属于常模参照性测评;飞行员的选拔与录用,主要属于效标参照性测评。

按测评范围划分,可分为单项测评和综合测评。组织诊断与人员培训过程中的测评,一般是单项测评;而人才选拔与绩效考评中的测评,大多属于综合测评。

按照测评技术与手段划分,有定性测评、定量测评,以及包括模糊综合测评在内的中性测评;按测评主体划分,有自我测评、他人测评、个人测评、群体测评、上级测评、同级测评及下级测评;按测评时间划分,有日常测评、期中测评与期末测评、定期测评与不定期测评;按测评结果划分,有分数测评、评语测评、等级测评及符号测评;按测评目的与用途划分,有选拔性测评、配置性测评、开发性测评、诊断性素质测评及考核性素质测评。

此外,还可以按测评活动分为动态测评和静态测评;按测评客体分为领导干部测评、中层管理人员测评、一般人才测评,等等。下面主要介绍我国目前常用的按照测评目的与用途进行划分的测评。

#### 1. 选拔性测评

选拔性测评的目的是选拔优秀人才,在人力资源管理活动中经常使用。待遇优厚、工作舒适的职位,往往会吸引众多的求职者,尽管前期可以采取一定的形式直接剔除许多不合格的求职者,但最后仍然会存在许多可供选择的求职者,此时就需要实施选拔性测评,从中选出与职位要求最为接近的被试者。

选拔性测评的主要特点有:第一,强调测评的区分功能。即要把不同素质、不同水平的人区别开来,把最优秀的求职者与一般性的求职者区分开来,便于雇主录用;第二,测评标准的刚性最强。测评标准应该严格、精确、“一刀切”,一旦实施就不能更改,否则所选拔出的“优秀者”就难以取信于民;第三,测评过程特别强调客观性。明

显标志就是对测评方法的信度的追求，表现为对数量化与计算机化的追求；第四，测评指标具有选择性。选拔性测评的指标允许具有一定的选择性，以客观、便于操作与相关性为前提，甚至可以是一些表面看上去与测评标准“风马牛不相及”的测评指标；第五，结果体现为分数或等级。

选拔性测评操作与运用的基本原则是公平性、公正性、差异性、准确性及可比性。公平性原则要求整个人才测评过程对于每个被试者来说，有利性相对平等，不是对某些人特别有利而对其他人不利，这是保证选拔性测评结果被公众接受的前提之一。公正性原则要求整个人才测评过程对于每个被试者来说，要求都是一致的，不是对某些人特别严格而对其他人随便。要求测评人员按统一的标准进行客观的测评，这是保证人们认为选拔性测评的结果有效的前提。差异性原则要求人才测评既以差异为依据，又能够反映被试者素质的真实差异，这是保证选拔性测评的结果正确性的前提。准确性原则要求素质测评对被试者素质差异的评估尽可能精确，限于误差允许范围内，这是保证人们对选拔性测评的结果信任的前提。可比性原则要求人才测评对被试者素质测评的结果具有纵向的可比性。一般要求采取量化的形式，不但可比，而且可以与其他测评结果相加。这是保证选拔性测评的结果最后在选拔人员过程中发挥实际作用的前提。

## 2. 配置性测评

配置性测评以人事合理配置为目的，是人力资源管理中比较常见的测评类型。实践表明，每个工作职位对其任职者都有一定的基本要求，当任职者现有的素质符合职位要求时，个体的人力资源就能主动发挥，创造出高绩效。否则，个体的人力资源就处于被动状态，甚至费尽九牛二虎之力也无济于事，低能低效。

配置性测评具有针对性、客观性、严格性、准备性等特点，主要表现为：第一，测评目的的针对性。配置性测评的目的是根据所配置的（工作）职位要求，寻找合适的申请者。如针对甲、乙两职位的测评过程必定有所区别，针对甲职位的配置性测评结果也不能运用到乙职位的人事配置上去。第二，测评标准的客观性。配置性测评的标准，必须以职位的客观要求为标准，不能主观随意制定。第三，测评的严格性。严格性既体现在测评的标准上，又体现在测评活动的组织与实施中。例如，飞行员的驾驶工作，绝不能因为找不到合格的人员而降低要求，但这并不意味着标准越高越好，让一个大学毕业生从事门卫工作的业绩，远远不如高中毕业生。第四，测评的准备性。这主要体现在人力资源管理过程的开端性上。依据配置性测评结果所做的人事配置，只是保证工作效率的必要条件之一，是一种可能性，是一种准备。随着工作要求与人员素质的变化，配置后要进行适当的调配，不要“一配定终身”。

### 3. 开发性测评

开发性测评,也称勘探性素质测评,是一种以开发素质潜能与组织人力资源为目的的测评,主要为人力资源开发提供科学性与可行性的依据。人的素质具有可塑性与潜在性,有些人当前也许不具备某方面的素质,但他可能具有发展这方面素质的潜力。如何发现这些人的潜力呢?显然有必要实施开发性测评。在每个企事业单位中,存在着不同类型的人力资源,要明确人力资源的基本形态,就必须实施开发性测评。有的人专注于技术运用,有的人热心于技术革新,有的人擅长技术传播,这些人实际具备了不同的人力资源形态,应该对他们分别采取不同的开发策略,以最大限度地发挥他们的作用。对于“运用型”的人,应把他们培养为“生产冠军”;对于“革新型”的人,应让他们有机会接触更多的技术资料,并对他们的失败抱以宽容的态度,鼓励他们的创新精神;对于“传播型”的人,应让他们横向发展,允许职位轮换流动。

开发性测评的特点主要有:第一,开发性测评带有调查性,主要了解被试者总体素质结构中,哪些是优势素质,哪些是短缺素质,哪些是显性素质,哪些是潜在素质,哪些素质具有开发价值等;第二,开发性测评一般是与素质潜能开发或组织人力资源开发相配合而进行的,是为人才开发服务的;第三,开发性测评的主要目的在于通过测评激励与促进各种素质的和谐发展,以及进一步提高,而不在于评定有无哪种素质或某种素质好与不好。

### 4. 诊断性素质测评

诊断性素质测评以了解素质现状或组织诊断问题为目的。在企业的组织管理中,为什么相同的职位、政策与环境,不同任职者的绩效水平却有很大的差异?如果需要从人员素质测评方面查找原因,最好实施诊断性素质测评。

诊断性素质测评与其他测评类型相比,主要特点有:第一,测评内容要么十分精细,要么全面广泛。诊断性素质测评的目的是查找问题的原因,因此测评时就像医生诊断病情一样,任何细节也不放过,测评内容十分精细与深入。如果是了解现状,则其测评的内容就十分广泛。第二,诊断性素质测评的过程是“寻根究底”的。一般是由现象观察出发,层层深入分析,直到找到答案。第三,测评结果不公开。其他类型的素质测评,结果一般都向有关人员公开,而诊断性素质测评的结果只供内部参考。第四,测评具有较强的系统性。诊断性素质测评要求从表面特征与标志观察入手,继而深入分析问题与原因,诊断“症状”,由此提出改进的方案,其他类型的素质测评无此要求。

## 5. 考核性素质测评

考核性素质测评又称鉴定性测评,是以鉴定与验证某种素质是否具备或者具备程度大小为目的的素质测评。考核性素质测评经常穿插在选拔性测评与配置性测评中。

考核性素质测评的特点有:第一,测评结果主要为被试者素质结构与水平提供依据或证明,是对被试者素质结构与水平的鉴定,而其他类型的测评结果并非如此;第二,侧重于被试者现有素质的价值与功用,比较注重素质的现有差异而不是素质发展的原有基础或者发展过程的差异;第三,具有概括性的特点。测评的范围比较广泛,涉及素质表现的各个方面,是一种总结性的测评,而其他类型的测评更明显的特点是具体性;第四,要求测评结果具有较高的信度与效度。

---

## 1.2 人才测评发展状况

---

人才测评自古已有,早在我国古代就已经蕴含了大量的人才测评思想。西方的学者则将人才测评发展为科学的理论与方法。人才测评的概念源于对人的心理特征和能力的测量,随着测评工作更加专业化,人才测评方法和人才测评工具也更加丰富,使得对人才的测评更加客观准确。目前,人才测评已经在各行各业得到广泛应用。

### 1.2.1 人才测评的起源

人才测评体系的起源可以追溯到 19 世纪初心理测量的诞生,它以 1905 年法国心理学家比奈与西蒙合作发表的智力测验量表——比奈-西蒙智力量表为标志,至今已有百年的历史。

#### 1. 比率智商

智力 (Intelligence) 是指生物 (或人) 一般性的精神能力,指生物认识、理解客观事物并运用知识、经验等解决问题的能力,包括记忆、观察、想象、思考和判断等。

比奈认为,智力包括三个方面的能力:一是选择并维持某一明确方向的能力;二是做出必要的调节,即策略调节以达到某种目标的能力;三是自我监控,即保证能做出必要策略调节的能力。

比奈和西蒙提出了“比率智商”(Intelligence Quotient, IQ),它是用来反映智力高低的高数,是心理年龄 (Mental Age, MA) 与实足年龄 (Chronological Age, CA) 的比值乘以 100。其计算公式为:

$$IQ = MA / CA \times 100$$



该公式的含义是：如果心理年龄与实足年龄相等，其智商为 100；如果心理年龄高于实足年龄，其智商较高（大于 100）；如果心理年龄低于实足年龄，其智商低于均值（低于 100）。

## 2. 比奈-西蒙智力量表的发展

（1）1905 年，比奈和助手西蒙编制了第一个心理取向的智力测验——比奈-西蒙智力量表，它有 30 个由易到难排列的项目，其中既有对较低级的知觉方面的测量，也有对较高级的判断、推理和理解等方面的测量。

（2）1908 年，比奈-西蒙智力量表进行了首次修订，修订后的量表运用了近代测验理论的基本思想，即测验的原理在于将个人的行为与他人比较并归类，首次采用智力年龄作为衡量儿童智力发展水平的指标。修订后的比奈-西蒙智力量表测验项目总数达到 54 个，并把测验题目按年龄分组，从 3 岁到 15 岁，只要每个年龄组的儿童中有半数人能通过的题目即属于这个年龄组的题目。

（3）1911 年，修订版的比奈-西蒙智力量表发布。这次修订没有重大变化，只是改变了几种年龄水平分组，并扩展到成人组。

（4）1916 年，斯坦福-比奈智力量表产生。随着比奈-西蒙智力量表被翻译成多种语言，各国心理学家对其进行了修订。其中以美国斯坦福大学刘易斯·麦迪逊·推孟主持修订的斯坦福-比奈智力量表最为著名。它的第一个修订本在 1916 年问世，这个修订本把比奈-西蒙智力量表从 54 项增加至 90 项，并对许多项目进行了修改，重新调整了年龄水平。该量表根据美国近 1 000 个儿童和 400 个成人组成的样组的结果重新加以标准化，不但对每个测试题的实施程序及评分方法做出了详细的说明和规定，更重要的是第一次将智商的概念运用到智力测验中，使智力分数能在不同年龄间比较，从而进一步发展和完善了比奈以智龄评定智力的方法。

（5）斯坦福-比奈智力量表的第二个修订本发表于 1937 年，其中有两套等值的测验，即 L 型和 M 型。它们各包含 129 个项目，内容涉及各种智力活动，尤为突出的是类比、对比、理解、词汇、异同、图画或语言填充、提出谬误、描画图样，以及记忆有意义的材料和数字等。

（6）斯坦福-比奈智力量表的第三个修订本发表于 1960 年。它是从 1937 年的 L 型和 M 型两套测验中选取最佳测试题而组成的，称为 L-M 型。这项工作以 1950 年至 1954 年 4 498 名 2.5~18 岁的被试者测验 L 型或 M 型中的测试题所得到的资料为依据。

## 3. 中国比奈量表

中国心理学家陆志伟曾两次主持修订比奈-西蒙智力量表。1924 年公布的《订正比

奈-西蒙智力测验说明书》仅适用于江浙一带儿童，所以陆志伟与吴天敏于 1936 年进行了第二次修订，使之也可适用于北方儿童。陆志伟的修订本的测验内容以刘易斯·麦迪逊·推孟的 1916 年的修订本为蓝本，共包括 54 个项目，其中某些项目按照不同要求应用于几个年龄组，故可按 75 项使用。项目性质大约可分为语言文字、数目、解图和技巧 4 类。3~11 岁的每个年龄组有 6 个项目，每项代表 2 个月；12 岁和 13 岁两个年龄组各为 3 个项目，每项代表 4 个月；14 和 15 岁合为 1 组共 6 个项目，每项代表 4 个月；16、17、18 岁合为 1 组共 9 个项目，每项代表 4 个月。其计分方法有两种，可按积分计算，也可按智龄分数计算，然后按智商公式算出智商分数。陆志伟自称，该测验用于 6~14 岁儿童较为可靠，6 岁以下或 14 岁以上儿童虽然能采用，但不及上述年龄范围准确。

中国的比奈-西蒙智力量表第三次修订本由吴天敏主持，称为《中国比奈测验》。它可用于 2~18 岁的被试者（每岁 3 题，共 51 题），最佳适用年龄是小学至初中阶段。评定成绩的方法采取离差智商。另有一个由 8 个试题组成的《中国比奈测验简编》，用起来较为方便、省时，试题包括对比、判断、数数、说反义词及情景问题等。例如，试题 1 是让被试者看 3 个大小不等的圆形，说出哪个大、哪个小。试题 2 为说出物名：爸爸写字用什么？妈妈切菜用什么？缝衣用什么？试题 7 为数手指：你这只手有几个手指？那手有几个手指？两只手共有多少个手指？试题 10 为看图说话：有 3 幅画，第 1 幅为 1 位小朋友在柜台前买东西的情景；第 2 幅为 1 位小朋友和爸爸妈妈一起逛公园的情景；第 3 幅为 1 位小朋友在车站给妈妈送雨伞的情景。让被试者看这些图画，并说出画的是什么。试题 14 为说反义词：快（慢）、甜（苦）、热（冷）、少（多）。一般来说，每道试题中的问题全部答对才算通过，如果连续 5 道题没有通过即可终止测定。

#### 4. 离差智商

离差智商的原理是用各个年龄段内人的智商的平均数作为参照，以被试者在同龄组中的分数为基础，看其分数与平均数的距离有多少标准差，从而确定其智商的高低。美国心理测量学家韦克斯勒在 1939 年编制了韦克斯勒成人智力量表，并提出离差智商的概念。

例如，两个年龄不同的成年人，一个人的智力测验得分高于同龄组分数的平均值，另一个人的智力测验得分低于同龄组分数的平均值，那么，就得出前者的 IQ 比后者高的结论。

## 1.2.2 人才测评的兴起与发展

人才测评的发展经历了从心理测验、能力测验、性格测验、职业心理测验发展到价值观测评和态度测评的转变。

### 1. 心理测验

心理测验起源于实验心理学中个别差异研究的需要。1879年,德国心理学家威廉·冯特在德国莱比锡大学设立了第一所心理学实验室,他在实验中发现,个体的行为相互间存在个别差异,个别差异的确定引起了心理测验的需要。心理测验的发展大约可以分为萌芽时期、成熟时期、昌盛时期及完善发展时期。

1869—1904年,心理测验处于萌芽时期。这一时期,心理测验尚未形成自己的体系,依附于实验心理学中个别差异的研究而存在,测验的内容大都限于感觉、动作或简单反应时间的测量,属于简单身体素质测评。

1883年,英国优生学家弗朗西斯·高尔顿在《人类才能及其发展的研究》一书中首先提出了“测验”这个术语,1884年他创设了“人类学测量实验室”,6年间测量了9337个人的身高、体重、视力、听力、色觉等素质特征。1890年,美国心理学家雷蒙德·卡特尔发表了题为“心理测验与测量”的论文,介绍了他编制的第一套心理测验,这套测验共10个题目,主要测量个体的感觉能力与动作过程。1894年,雷蒙德·卡特尔首先用这种测验方法测量哥伦比亚大学的学生,使测验走出实验室,直接应用于实际。

1905—1915年,心理测验发展到成熟时期。19世纪后期,威廉·冯特、弗朗西斯·高尔顿、雷蒙德·卡特尔都对感觉能力的测量进行了研究。然而对心理测验的发展有最重要影响的应数法国的比奈,他与西蒙于1905年合作编制的比奈-西蒙智力量表,是世界上第一个标准化的心理测验。

1916—1940年,心理测验处于昌盛时期。这一时期,不仅智力测验在广度与深度上有了突破性的发展,而且一般能力测验、特殊能力测验、人格测验相继出现,心理测验在理论上得到完善,在应用上得到了空前的发展。测验的表现形式,由文字扩展到图形、操作等非文字,由直接的测量扩展到投射与预测的测验,测验的功用由研究走向社会服务。

1937年,雷蒙德·卡特尔编制了“16PF”。1938年,瑟斯顿提出了“主要的心理能力论”,概括出7种主要的智力:知觉速度、推理能力、言语理解能力、语词的流畅性、空间知觉、记忆及计算能力。同年,亨利·默里与摩根编制了投射测验之一的主题统觉测验。

1941年至今,心理测验处于完善发展时期。在这一时期,心理测验一方面接受“教

育评价运动”的“挑战”；另一方面在测验的理论、技术与编制方法上，都有非常大的进步。20世纪40年代初，哈瑟韦和麦金力编制了调查个人适应性和社会适应性的MMPI；1953年，艾森克夫妇编制了EPQ。这一时期，还兴起了职业性向与职业技能测验的新高潮，用于挑选各行业的职员与管理人员。

从心理测验的起源与发展可知，心理测验产生于对个别差异鉴别的需要，它广泛应用于教育、企事业人员的挑选与评价。在这一过程中，人们编制了许多心理测验。其中，比较有影响力的心理测验包括比奈-西蒙智力测验（1905—1911年）、斯坦福-比奈儿童智力测验（1916年）、罗夏墨迹测验、亨利·默里与摩根的主题统觉测验、16PF、皮亚杰故事测验、科尔伯格两难故事测验、雷斯特确定问题测验等。在所有心理测验的定义中，阿纳斯塔西所下的定义比较确切：心理测验实质上是行为样组客观的和标准化的测量。这个定义告诉我们以下几点。

（1）心理测验是对行为的测量。这些行为主要是心理的而不是反射性的生理行为，是外显行为而不是内部心理活动，是一组行为而不是单个行为。

（2）心理测验是对一组行为样本的测量，即所测量的行为组是有代表性的一组行为。任何个体在不同时间、空间与条件下的行为表现是不尽相同的，如果我们所测评的行为抽样不同，得到的结果就会不同。

（3）心理测验和行为样组不一定是真实行为，而往往是概括化了的模拟行为。例如，罗夏墨迹测验，答题行为均不是真实的行为，而是一种间接的行为反应。

（4）心理测验是一种标准化的测验。所谓标准化，在这里是指测验的编制、实施、计分及测验分数解释程序的一致性，这是测验的内在要求。如果想使测验的最后结果具有可比性，测验的条件就必须具有等同性或统一性。

（5）心理测验是一种力求客观化的测量。

从上述所有测验中可以看出，这些测验所采用的技术，如采用问答、填空、选择等客观性试题，都是尽可能排除人为的主观影响。需要注意的是，这些测验并不是完全客观化的。

## 2. 能力测验

能力测验是用来测定被试者成功完成某些活动所必须具备的个性心理特征。按能力种类划分，能力测验可以分为一般能力测验和特殊能力测验。

一般能力是指大多数活动所共同需要的能力，是人们所共有的基本能力、普通能力，如观察力、注意力、思维力、想象力等，就是通常所说的智力。特殊能力是指为完成某项专门活动所必需的能力，其只在特殊活动领域内发挥作用，如机械运动能力、数学能

力、音乐能力、写作能力等，都是完成相关活动必不可少的能力。

### (1) 智力测验。

智力测验是通过测验方法来衡量人的智力水平高低的一种科学方法,由于人们常把智力看作各种基本能力的综合,所以智力测验又可称为普通能力测验。

韦克斯勒成人智力量表是常用的一种智力测验,可以用来测量 16~74 岁人的智力水平。此量表包括言语量表和操作量表,分别包括 6 个和 5 个分测验(见表 1-2),这些分测验从各个层面来测量智力。在施测时,按言语和操作两个方面交替进行。韦克斯勒成人智力量表每次只能测验 1 名被试者,每测 1 名被试者需要用 1.5~2 小时,由于是单独施测,所以人力、财力成本较高,在选拔中并不常用。但是它在管理能力上有较好的效度,所以在选拔高层管理人员时比较有效。在使用此量表时,需要由训练有素、有丰富经验的管理学家担任测评人员,以保证评分和分数解释的准确性。

表 1-2 韦克斯勒成人智力量表的测验项目和内容

测验项目	内 容
言语量表	
常识(一系列有关人们日常生活的开放式问题)	知识广度
理解(一系列有关人们对社会规则了解情况的开放式问题)	实际知识与理解力
心算(所有问题是难题形式,以解决的正确性及所需时间计分)	算术推理能力
类同(阅读配对名词,找出其共有的特质)	抽象能力与鉴别能力
背诵数字(测评人员读出数位数字,由被试者依记忆重复)	注意力与记忆力
词汇(被试者对所给难度逐渐增加的字加以定义)	语言知识与表达能力
操作量表	
译码(一连串符号均各有一个数字为其代码,在限定的时间内,看着符号写出其正确的数字代码)	学习与书写的速度
填图(指出每张图片中遗失缺漏的部分)	视觉记忆力
积木图案(将积木尽快依图示拼凑成与图式相同的形象)	视觉的模式分析能力
图片排列(尽快将图片部分拼成一副完整的图画)	对社会情景的理解力
图像组合(在限定的时间内将图像部分拼凑完整)	处理部分与整体关系的能力

瑞文测验也是常用的一种智力测验。它是一种非文字测验,试题是一系列彩色或黑白的图形,右下角缺了一块,要求被试者从这些图形的变化中找出规律,然后在提供的 6 小块或 8 小块图形中挑选 1 块最合适的图形填补上去,使图形完整。瑞文测验是一种团体测验量表,因此使用较为方便。评分时正确得 1 分,错误得 0 分,然后按总分及被

试者的年龄即可在常模表中查到其智商分数。瑞文测验具有一般文字智力测验所没有的特殊功能,由于是用图形方式出题的,其对语言有障碍的被试者、少数民族被试者都适用,可以看作一种公平的测试。由于其题量少、评分简单,所以施测省时省力,是进行大规模智力筛选和考查的理想工具。

### (2) 特殊能力测验。

特殊能力反映人们在某些方面的特长和技能,通过对这些特殊能力的测量可以推论、评估人们的较高层次的能力。有许多职位对任职者是否具有某方面的能力会提出一定的要求,这时运用特殊能力测验就可以为选拔提供参考依据。特殊能力测验包括机械能力测验、文书能力测验、音乐美术能力测验等。贝内特机械理解测验就是机械能力测验的一种,它可以测量被试者对机械关系和物理定律的理解能力。它的测验材料是以图形加问题的形式,要求被试者根据图形回答问题。机械能力测验可用来衡量人们对机械原理的理解、判断空间形象的速度、准确性,以及手眼协调的运动能力。该测验应用范围较广,有效对象为机械工程师、修理工、工匠和工程师等职位的应聘者。文书能力测验包括与智力测验类似的题目,以及测量知觉速度和准确性的题目,因为文书工作需要言语、数学能力和动作敏捷、察觉异同的敏锐力等,所以文书能力测验主要用于测验个人打字、速记、簿记、处理文书和联系工作的能力。常用的文书能力测验包括测量被试者的文书速度、准确性、言语准确性和数字能力的一般测验,以及简单数字和姓名匹配检查的明尼苏达文书测验。

### (3) 多重能力倾向测验。

测量人的各种能力,可以为每种特殊能力编制单独的测验,但这样往往比较费时,在人事选拔中也难以实现,因此有必要使用能同时测定几种能力的测验,多重能力倾向测验就可以满足这一要求。能力倾向是一种潜在的能力素质,即有效承担某项工作或成功履行某种职责的可能性,而并非已经具有的现实智能水平。它表现为一个人在一定环境中经过适当学习训练,能够做什么的潜力,即个人能学会做什么。能力是由先天因素、后天学习和社会实践而形成的完成某项工作任务或活动的智能水平,或者说从事某种活动(如承担某项责任、完成某项任务等)现已具有的主观条件,而不需要进一步训练。因此,能力倾向既不同于智力,也不同于人在某方面接受教育和训练获得的专业知识和技能,它具有相对广泛性,可以影响一个人在各种领域中多种活动的效率。能力倾向还具有相对稳定性,它是由个人多年学习和实践积累形成的、稳定的心理特征,很难改变,虽有可能改变,但不可能在短时期内迅速提高。

多重能力倾向测验就是用来测量与某些活动有关的一系列心理潜能的考试,能同时测定多种能力倾向。在各种人员的选拔方案中,多重能力倾向测验都被认为是必不可少

的部分。这种测验主要有基本心理能力测验（Primary Mental Abilities Test, PMAT）、普通能力倾向成套测验（General Aptitude test battery, GATB），以及区分能力倾向测验（Differential Aptitude Test, DAT）等。其中，GATB由8个纸笔测验和4个仪器测验组成，可以测量9个因素：言语能力、数字能力、空间判断能力、一般学习能力、形状知觉、文书知觉、运动协调能力、手指灵巧度、手部灵活度。这9个因素中的不同因素组合代表着不同的职业能力倾向，如数字能力、空间判断能力较强，手部灵活度较好的人适合从事设计、制图作业。因此，GATB也常用来测定被试者的职业倾向。

#### （4）行政职业能力倾向测验（Administrative Aptitude Test, AAT）。

AAT是国家公务员录用考试公共科目的重要组成部分。它有利于帮助人事部门了解被试者从事行政工作的潜能差异，避免选人过程中出现“高分低能”的现象。AAT成绩的优劣，将在很大程度上决定被试者能否加入公务员队伍。

AAT属于一种潜能测试，即考查被试者在未来从事公务活动中潜在的能力素质，考查被试者有无成功履行公务员职责的可能性。因此，它既不是一种理论知识的考试，也不是行政职业技能测验，而是介于二者之间。通过AAT只能说明被试者具备了进入国家机关工作的必要条件，即AAT只是在一定程度上体现了对公务员最低限度的要求，还需要综合其他测评结果的考虑才能决定录用与否。

在内容结构和表现形式上，AAT有测验面宽、题量大、题型稳定、题本与答卷分离等特点。该测验通常包括以下5个相对独立的分项测验，共150题。

① 知觉速度与准确性测验。考查被试者对数字、字母等视觉符号快速和准确地感觉、比较、转换、加工的能力，涉及感觉、知觉、短时记忆、判断等心理过程，是一种典型的的速度测验。这部分测验单独计时，答错要倒扣分。

② 数量关系理解测验。测验题目所用的数学知识都很简单，主要考查被试者的算术能力。

③ 言语理解与表达测验。考查被试者对文字材料的理解、分析和运用能力。

④ 资料分析测验。以图形、表格文字表达的统计资料往往包含大量信息，如何准确理解、综合分析这些信息，是公务员必须具备的重要能力。

⑤ 判断推理实验。该实验涉及对图形、词语概念、事件关系和文字材料的理解、比较、演绎及综合判断能力，判断推理能力的强弱反映一个人对事物本质及事物间联系的认知能力的高低。公务员必须具备较强的判断推理能力。

AAT的测试内容主要以文字、数表和图形3种形式表现，一律采用客观性的标准化试题。与一般的标准化考试不同，AAT有较严格的时限要求，其中，知觉速度与准确性测验的每道试题平均作答时间仅为10秒钟。在规定时间内不能完成所有题目是常

有的现象。因此，AAT 的作答要求被试者思维反应敏捷，能迅速把握题意，并马上做出反应。

在测验实施上，AAT 也较为严格。由于有较多的时间限制，因此要求程序严密、操作规范。测评人员应严格按施测程序操作，并控制被试者按指令作答，才能保证测试的有序进行。

对被试者来说，该测验测试的是个人由多年生活、学习和实践积累形成的稳定的心理特征。因此，不能靠短时突击的死记硬背来提高成绩。其实，AAT 测试题目并不难，如果有足够时间的话，多数人都可以取得较好的成绩。但若在时间有限、题量又大的紧张状态下，有些考生就难以正常发挥了。因此，它不仅是被试者潜在能力水平的竞争，也是被试者个体胆识、意志力、信念、适应性等心理品质经受全面考验的过程。因此，只有以自信、乐观的心态接受考试，严格按照测评人员的指令答题，才能正常发挥，取得客观、真实的成绩。

#### （5）创造力测验。

创造力是指产生新思想、发现和创造新事物的能力，是更高级的思维层次。在人员招聘选拔中，几乎所有的组织都倾向于选拔有创造力的员工，创造力往往会给组织带来意想不到的收益。因此，在心理测评中，创造力测验是非常重要的衡量创造性的工具。

主要的创造力测验有南加利福尼亚大学发散性思维测验、托兰斯创造思维测验及芝加哥大学创造力测验等。以南加利福尼亚大学发散性思维测验为例，它主要以发散性思维（一项重要的创造思维）为测量内容，由 14 项测验组成，分别要求被试者对言语和图形内容做出反应，如迅速写出含有字母“O”的单词（词语流畅性）；列举如果人不睡觉可能产生的后果（对事件后果的估计）；在给定的图形上增加线条，使之成为可辨认的物体（绘图）；移动指定数目的火柴以形成特定数目的正方形或三角形（火柴问题），等等。该测验适用于中学文化水平以上的人，具有良好的信度。现有的这些创造力测验多属于开放性题目，因此评分往往带有较大的主观性。在使用这一类创造力测验时，应严格参照评分标准进行评分，并且使用经过严格训练的测评人员，这样才能保证测验结果的可靠性和有效性。

### 3. 性格测验

常用的性格测验主要有以下几种。

#### （1）MMPI。

MMPI 是问卷法的一种，它可以同时测量多种人格特质，因此被称为“多相”测验。它由 566 题组成，其中 16 题重复出现，以检验被试者反应的一致性和做题的态度。被



试者对每个问题根据自己的实际情况选择“是(T)”“否(F)”两个中的一个,一般耗时 1 个小时左右。MMPI 有 10 个临床量表,可以得到代表 10 种人格特质的分数,如表 1-3 所示。

表 1-3 MMPI 量表所示人格特质

序 号	临床量表	符 号	高分的解释
1	疑病	Hs	强调身体的疾病
2	抑郁	D	不快乐、抑郁
3	歇斯底里	Hy	对应少许的反应是否有问题
4	精神病态	Pd	与社会缺乏一致性,经常处于法律纠纷中
5	男性化—女性化	Mf	女性倾向(男子);男性倾向(女子)
6	妄想性	Pa	多疑
7	精神衰弱	Pt	烦恼、焦虑
8	精神分裂症	Sc	孤独、古怪思想
9	轻躁症	Ma	冲动、激动
10	社会内向	Si	内向、害羞

MMPI 还包括四个与效度有关的量表:疑问量表、说谎量表、效度量表、校正量表。它可以用来考查被试者的作答是否可信、态度是否认真、是否有较强的防御机制、是否在回避问题,以及是否在伪装疾病等。首先,通过被试者的回答得到其原始分数;其次,将原始分数转化为标准分(对照常模表),标注在剖面图上;再次,将各量表标准分的点按顺序连接起来,就得到被试者的 MMPI 剖面图;最后,通过对剖面图形态的研究,可以对被试者的多相性格做出合理解释。

MMPI 的优点是题目较多,能够全面地测量人格特质的各个方面,信度、效度较高。但题目多,会影响被试者做题时的情绪,并且它对各种人格特质的描述多用病理上的名词,用在正常人身上不是很合适。一般来说,MMPI 的施测和评分并不困难,但在对剖面图的分析上,必须是有丰富临床经验的心理学家才能够承担。所以在实际运用中,MMPI 还有一定的局限性。

## (2) 16PF。

16PF 适用于 16 岁以上的青年和成年人,它可以在约 45 分钟里测量互相独立的 16 种人格特质(见表 1-4)。

表 1-4 16PF 中的 16 种人格特质

因素	特质名称	低分者特性	高分者特性
A	乐群性	缄默、孤独	外向、热情
B	智慧性	迟钝、学识浅薄	聪慧、富有才识
C	稳定性	情绪激动	情绪稳定
E	好强性	谦逊顺从	好强固执
F	乐观性	严肃审慎	轻松兴奋
G	有恒性	权宜敷衍	有恒负责
H	敢为性	畏怯退缩	冒险敢为
I	敏感性	理智、注重实际	敏感、感情用事
L	怀疑性	信赖随和	怀疑、刚愎
M	幻想性	现实、合乎成规	幻想、狂妄不羁
N	世故性	坦白直率、天真	精明能干、世故
O	忧虑性	安详沉着、有自信心	忧虑抑郁、烦恼多
Q1	实验性	保守、传统	自由、批评、激进
Q2	独立性	依赖、附和	自立、当机立断
Q3	控制性	矛盾冲突、不明大体	知己知彼、自律严谨
Q4	紧张性	心平气和	紧张

16PF 根据实验统计结果所得的公式推算出许多种可以形容人格类型的次级因素（见表 1-5）。这些次级因素不是直接用原始分推算的，而是由几个有关的基本因素的标准分，经过数量权衡，连同指定数相加而成的。常数和均衡数量的多少是由雷蒙德·卡特尔根据多年的研究统计分析结果获得的。

表 1-5 16PF 测定的人格类型

双重个性因素	高分者特性	低分者特性
适应与焦虑性	焦虑性高、易激动、易出现不满意感	焦虑性低、可能对困难的工作缺乏毅力
内向与外向性	外向、开朗、善交际、不受拘束	内向、趋于胆小、自足，和别人交往时拘谨、克制
感情用事与安详机警性	安详机警、富于事业心、果断、刚毅，但也易贸然行事、不顾后果	感情用事、情感丰富、含蓄敏感、性格温和、讲究生活艺术

续表

双重个性因素	高分者特性	低分者特性
怯懦与果断性	果断、独立、有气魄、有攻击性、有创造性	怯懦、顺从、被动、依赖、迎合迁就别人

此外,雷蒙德·卡特尔还用因素分析法演绎出了估量各种行为及其有关人格因素分数高低标准的特殊公式,包括评定心理健康者的个性特质公式、创造性强者的个性特质公式、专业而有成就者的个性特质公式,以及在新环境中成长能力的个性特质公式等。这些公式可以作为心理咨询或升学指导的诊断工具。

16PF 具有较高的信度和效度,编制比较科学,施测、评分和解释都比较方便。因此,16PF 是国际上颇有影响力的一种人格测验,得到了非常广泛的运用。

### (3) 加利福尼亚心理测验 (California Psychological Inventory, CPI)。

CPI 是由美国加利福尼亚大学心理学教授高夫设计的一种个性测验。该测验有一半题目来自 MMPI,但与 MMPI 不同的是,CPI 更强调人的正常的个性特质。CPI 由 480 道题组成,可分成如表 1-6 所示的 4 个量表群,共 18 个分量表,每个分量表都可以评估一种在社会交往中表现出的个性特征。

表 1-6 CPI 的量表群及其分量表

量表及所测因素		分量表	内 容
第 1 量表群	人际关系 社会适应性	支配性	领导能力、支配性和主动性
		上进心	个人争取达到某种社会地位的能力和技术的程度
		社交性	个人参与社交和人际关系的程度
		自在性	在社交时的自在性和自信的程度
		自尊性	自我价值、自我接纳、独立思考与行动的程度
		幸福感	身心健康和自我满意的程度
第 2 量表群	社会化 成熟度 责任心 价值观	责任心	责任心和诚实的程度
		社会化	社会成熟程度和正直性程度
		自制力	自我控制的程度
		容忍性	对人对己的宽容程度
		好印象	社会认可性和可靠程度
		从众性	个人与一般人符合的程度

续表

量表及所测因素		分量表	内 容
第 3 量表群	成就潜能 智力水平	遵从成就	对成就关心或要求的程度
		独立成就	对社会自主及促进其成就的关心程度
		智力水平	智力程度
第 4 量表群	生活态度和倾向	共鸣性	一个人对他人的需要、动机和兴趣的敏感程度
		灵活性	个人变通能力
		女性化	个人的女性化程度

CPI 适用于 13 岁以上的正常人，可以个别施测，也可以团体施测，因此该量表使用很方便，应用也非常广泛，尤其是在企业职业选拔中表现出较高的实用价值。该量表适用于正常人的个性评定，除了测量被试者现在的性格特征，还可以预测被试者今后的学业成绩、犯罪倾向和职业成功的可能性等。如果被试者几乎所有的分数都超过平均标准分数线，则表明他可能在社交和智力两方面都发展得较好。用 CPI 可以测出行为异常、情绪不稳定者，减少这些人对组织发展的不良影响，确保录取人员行为和情绪属于正常范围，并根据各岗位的职责要求选拔个性特点适宜的员工。

#### （4）投射测验。

投射测验是间接测量人格的一种方法，它使用投射技术唤醒被试者的内心世界或个性的不同表现形式，从而根据被试者的反应分析和挖掘其内在的需要等心理状态。

投射测验使用含糊、模棱两可的刺激材料，让被试者在不受限制的情况下，自由做出反应，不知不觉地表露出自己的人格特征。在投射测验中，被试者一般不清楚测验的目的，因此可以较大程度地避免“防卫机制”的出现。与问卷法不同，投射测验主要关注对性格的总体评估，而不是个性因素和逐个描绘，它让被试者通过对不明确情景的自然反应，使被试者潜意识中的欲望、冲突得以流露，从而整体地研究其人格特征。因此，在问卷法中难以触及的潜意识，在投射测验中很可能表露出来。

罗夏墨迹测验是一种比较经典的投射测验。它向被试者逐一展示 10 张没有内容、没有意义的墨迹图，然后问被试者：“你看到什么？”“它像什么东西？”“它让你想到什么？”等，让被试者完全根据自己的联想和感受自由回答。测评人员则把被试者的反应都详细地记录下来，根据计分规则评分并解释测验结果。然而，罗夏墨迹测验的评分和解释是一项极为费时费力的工作，做起来非常困难，只有训练有素且临床经验非常丰富的心理学家才能掌握和熟练运用这种方法。此外，罗夏墨迹测验需要个别进行，不适宜进行大规模的使用。

主题统觉测验也是一种著名的投射测验,它的材料是30张印有人物、风景的图片,以及1张空白卡片,对不同的被试者选用不同的图片施测。被试者的任务是根据不同的图片自由编造一个有戏剧性的故事,测评人员则记录被试者的叙述及各种其他反应,也可以要求被试者把故事写下来。测评人员在测验后还应与被试者谈话,以便深入了解故事的内容。由于个人对图画所编造的故事与他的生活经验有密切关系,因此在编造故事时,人们常常不自觉地把个人的心理活动和潜意识投射到故事中去。通过分析故事,可以揭示被试者的一些主要的动机、情绪和个性等,主题统觉测验的特殊价值就在于它不但展示个体潜在的、被压抑的倾向,还能整体地描述个性。

投射测验是以被试者的自由想象和发挥为中介,从而推论被试者内心潜伏的心理欲望、心理冲突等总体的人格特征,并可以同时考查个人的智能、创造力、解决问题的能力。它对测验结果的评分和解释有非常大的主观性,对测评人员要求也非常高,是一种比较费时、费力的测验方法。但它是对问卷法测验的补充,可以揭示人格中不易显露的一面。因此,该测验虽然不能广泛地运用于人事考评中,但具有一定的实用价值。例如,可以适当地用于评价高级总裁职位的候选人,作为选拔决策的重要依据。

#### 4. 职业心理测验

了解被试者的职业心理因素所属的类型及特点,即职业倾向、职业兴趣等,无论对于被试者确定合适的求职目标,还是对于预测被试者在某职业中的可能成就和发展以作为选拔依据,都是非常重要的前提。要把握个体的职业心理特性,可以凭经验做出判断,但为了使评估更客观和有效,还是应使用职业心理测验量表。

##### (1) 霍兰德职业倾向测试量表。

霍兰德职业倾向测试量表是常用的职业心理量表之一,它是由美国职业指导专家霍兰德根据他本人大量的职业咨询经验,以及他所创立的“人格类型”理论的基础上编制的测评工具。霍兰德在其一系列关于人格与职业关系假设的基础上,提出了6种基本的职业类型,并解释了这6种职业类型基本的人格倾向,如表1-7所示。

表 1-7 霍兰德职业类型与基本的人格倾向

职业类型	基本的人格倾向
实际型	喜欢有规则的具体劳动和需要基本操作技能的工作,但缺乏社交能力,不适应社会性质的职业。其典型的职业包括技能性职业(如一般劳工、技工、修理工等)和技术性职业(如软件开发师、摄影师、机械装配工等)
研究型	具有聪明、理性、精确等人格特性,喜欢抽象的、分析的、独立的定向任务这类研究性质的职业,但缺乏领导才能。其典型的职业包括科学研究人员、工程师等

续表

职业类型	基本的人格倾向
艺术型	具有想象、冲动、直觉、理想化、有创意、不重实际等人格特性，不善于事务性工作。其典型的职业包括艺术方面的（如演员、导演）和文学方面的（如诗人、剧作家等）
社会型	具有合作、友善、善社交、善言谈、洞察力强等人格特性，喜欢社会交往，关心社会问题，有教导别人的能力。其典型的职业包括教育工作者与社会工作者
企业型	具有冒险、有野心、独断、乐观、自信、精力充沛等人格特性，喜欢从事领导及企业性质的职业。其典型的职业包括政府官员、企业领导等
传统型	具有顺从、谨慎、保守、实际、稳重等人格特征，喜欢有系统、有条理的工作。其典型的职业包括办公室人员、会计、打字员等

霍兰德职业倾向测试量表通过让被试者在一系列工作中做出符合自己喜好意愿的选择的方式，统计分析所有被选的答案，来发现和确定被试者的职业兴趣和能特长。这一量表共由 7 个部分组成。

第 1 部分：你心目中的理想职业（专业）。要求填写最想从事的 3 种工作或最想读的 3 种专业。

第 2 部分：你所感兴趣的活动。在分归于 6 类的 60 种活动中，用选择“是”“否”的方式回答“你是否喜欢从事下列活动。”

第 3 部分：你所擅长或胜任的活动。在分归于 6 类的 60 种活动中，选出你能做或大概能做的事，在“是”栏里打“√”，反之，在“否”栏里打“×”。

第 4 部分：你所喜欢的职业。在分归于 6 类的 60 种职业中，选出自己感兴趣的职业。

第 5 部分：你的能力类型简评。评定自己在 6 类职业能力方面的大致水平。

第 6 部分：统计和确定你的职业倾向。将各部分中各类的测验分数（即选“是”的个数）填到得分表中，并相加得到总分。然后将 6 种职业倾向得分按高低排列。

第 7 部分：你所看重的东西——职业价值观。从所附的 9 项工作价值标准中选出你认为最重要和最不重要的两项因素。

测验完成后，可以对照职业索引表，判断被试者的职业兴趣，以及适合被试者的职业类型。这一量表使用非常简单，在职业指导和人才选拔中运用十分广泛。

## （2）气质量表。

气质是个人相对稳定的心理活动的动力特征，它使人的日常生活带有一定的色彩，

形成一定的风貌。构成气质类型的心理特征有感受性、耐受性、不随意反应性、反应的敏捷性与灵活性、可塑性、与稳定性、内外向性、情绪兴奋性、情绪和行为特征。这些不同的心理特征构成了4种不同的气质类型：胆汁质、多血质、黏液质和抑郁质。一般来说，具有某种典型气质特征的人是很少的，三种气质的混合型也很少，多数人是近似其中某一类型或者是两种类型的混合气质。气质与神经类型的强度、均衡性、灵活性、行为特点，如表1-8所示。

表1-8 气质与神经类型的强度、均衡性、灵活性、行为特点

气质与神经类型	强度	均衡性	灵活性	行为特点
胆汁质/兴奋型	强	不均衡	不灵活	攻击性强、易兴奋、不易约束、不可抑制
多血质/活泼型	强	均衡	灵活	活泼好动、反应灵活、好交际
黏液质/安静型	强	均衡	惰性	安静、坚定、迟缓有节制、不好交际
抑郁质/抑制型	弱		不灵活	胆小畏缩、消极、防御反应强

气质在人的实践活动中不起决定性作用，但有一定的影响，主要表现在它可能影响活动的效率。例如，要求做出迅速灵活反应的工作，具有多血质和胆汁质的人比较合适，而具有黏液质和抑郁质的人则较难胜任；反之，要求持久细致的工作，具有黏液质、抑郁质的人较为合适，而具有多血质、胆汁质的人又较难胜任。显然，为了提高工作效率，对不同职位和岗位的员工的气质特性就要提出特定的要求，有些特殊工种还有其特殊要求，否则是难以适应和胜任工作的。华东师范大学俞文钊教授研究发现，多血质、黏液质，或者两者的混合型是较适宜担任管理人员的气质类型。

常用的气质量表有瑟斯顿气质量表、吉尔福特-齐默尔气质调查表，以及陈会昌气质量表等。陈会昌气质量表共60道题，每种气质类型包括15道题。要求被试者对题目的陈述根据自己的实际情况在“很符合”“比较符合”“介于符合与不符合之间”“比较不符合”“完全不符合”中进行选择。采取五级评分制计分，根据得分多少确定气质类型。该调查表简单易行，信度和效度较高。

## 5. 价值观测评

价值观是个人对客观事物的意义与重要性的总评价，它使人的行为带有个人的某种稳定的倾向性，人们对各种事物的评价，在心目中有轻重主次之分，这种主次的排列，就构成了一个价值体系。价值观及其价值体系是决定个人行为与态度的基础。在相同的条件下，不同价值观的人会有不同的行为与态度，价值观不仅影响个人行为，还影响整个组织的行为，进而影响企事业单位的经济效益与社会效益。因此，在人员录用时，必须考虑价值观测评。

价值观的类型有不同的划分，其中有两种类型对测评有参考价值。

(1) 斯普兰格六类型。

① 理论型。理论型的人表现为乐于发现真理，凭借观察和推理发现事物之间的一致性和差异性，具有实验的、批判的和理性的爱好。

② 经济型。经济型的人强调事物的实用性，凡事以有效与实惠为尺度。

③ 审美型。审美型的人总是从形式与和谐中寻求最高价值。以文雅、优美、对称和恰当的标准判断每种经验。

④ 社会型。社会型的人的特征是利他与仁慈，他们在社会实际生活中往往表现出宽容、富于同情心和无私等品德。

⑤ 政治型。政治型的人热衷于追逐个人权力、影响力和声望等，他们在政治以外的活动领域中，也希望通过竞争、奋斗获得名誉与地位。

⑥ 宗教型。宗教型的人是一种理想信念主义者，他们在工作和生活中总是以自己的信仰与理想为准则。

(2) 格雷夫斯七等级型。

① 反应型。反应型的人只追求自己的基本生理需求获得满足，而不考虑其他人与周围条件。

② 宗教式忠诚型。宗教式忠诚型的人喜欢按部就班地看问题，依赖成性，服从于传统习惯与权势。

③ 自我中心型。自我中心型的人粗犷、富有闯劲儿，为了取得自己想要得到的东西，愿意做任何工作。

④ 顺从型。顺从型的人的特征是忠诚、尽职尽责、勤恳、谨小慎微、喜欢任务明确。

⑤ 权术型。权术型的人非常现实，追求功利，喜欢玩弄权术，乐于奉承有关的上级。

⑥ 社交中心型。社交中心型的人重视工作集体的和谐，喜欢平等的人际关系，把善于与人相处和被人喜爱看得重于自己的发展。

⑦ 价值主义型。价值主义型的人喜欢自由和创造性的工作，喜欢灵活的职务，重视挑战性的工作和学习成长的机会，把金钱和晋升看作次要的，能高度容忍不同观点的人和模糊不清的意见，对于僵化的制度、滥用权力等现象，能直言不讳。

价值观是一种高级的认知形式，问卷测验是常用的一种测评形式。这方面国外较为推崇的是阿尔波特、韦农和林达塞所编制的价值观研究量表，此量表发表于1931年，后来在1951年与1960年分别进行了修订。



价值观研究量表是以斯普兰格六类型为依据编制的，共 45 道题，前 30 题每题 2 个选项，后 15 题每题 4 个选项，下面是该量表中的两道题。

第一部分例题：如果你在报纸上看到同样大小标题的新闻，你倾向于先读（ ）。

- A. 反对派领导人协商解决双方冲突      B. 世界投资环境有明显改善

第二部分例题：根据你自己的意见，一个整年都忙碌工作的人，星期天最好（ ）。

- A. 读一些严肃的书籍                      B. 参加体育比赛  
C. 去听一场交响乐音乐会                D. 去听一次好的布道

所有的题目都由自己作答，不受时间限制，允许但不希望被试者有不作答的题目，对于第一部分试题的计分方法是对于 A、B 两个选项，若肯定一个而否定另一个，则分别赋予“3”分和“0”分；如果只是稍微偏向一个并不否定另一个，则分别赋予“2”分和“1”分。对于第二部分试题的计分方法是，给最赞同的选项“4”分，最不赞同的选项“1”分，中间依此类推。

该量表总共 120 个选项，其中对应每种类型的选项各 20 个，经过相加，获得 6 个总分。被试者在哪种类型上的总分最高，他就倾向于哪一种价值观类型。

## 6. 态度测评

态度是个体对某类型的人、事物、机构及观念的较为稳定的行为倾向，其定义是，个体针对一组社会事物而产生的具有一致性的反应群。它的表现形式有三个层次：一是认知层面，即个人对事物的了解、认识、看法，其中最重要的是评价性的见解；二是情感层面，即个体对事物的情感与好恶；三是行为倾向层面，即个体对事物的可观察的行为倾向。例如，怕苦怕累思想与情绪较严重的人不愿意到第一线工作。每个层面都有方向性（赞成与反对）、强弱（非常、比较、一般），以及多面性（指每一层面由很多种类构成，如爱可以分为友谊、吸引、尊敬、恋爱等）。三个层面之间还存在一致性。

意见与态度、价值观作为表达个人理性与情感的评价形式，其功能是相同的，但是它们又有不同的地方。意见通常指个人对特定的、具体的事物对象做短暂性的评价，大多涉及公共事物；态度所涉及的对象更广泛、更抽象些，持续性比意见更久；价值观也指信念，较态度更基本、更持久些。

由于态度主要是一种内心活动形式。在认知、情感与行为上具有一致性，因此对态度的测评，最常用的是问卷测验和认知量表形式。

下面结合几种态度量表介绍一下有关的编制方法，以了解态度测评的原理与方法。

### （1）瑟斯顿态度量表的编制方法。

- ① 根据所要测评的态度对象，收集与对象有关的态度语（叙述句）100~200 句。

收集态度语的原则是：措辞简单，语句通顺；语句意义易于了解，不产生歧义；每个态度语需要对态度对象表明一个确切的态度；公认的理论或确切的事实不适合用作态度语；每个态度语必须是一个完整的句子。

② 请 50~60 位对态度对象熟悉而又客观的专家，按每个态度语表示的态度将所有的态度语划分为若干组，一般是 7~13 组。请专家按照下面的要求把态度语归到相适宜的类别组中：第 1 组内放置态度最极端反对的态度语，第 2 组内放置那些最为赞成的态度语，其余各组按顺序放置，使第 2 组内的态度语尽管是反对的，但不如第 1 组那么强烈，第 3 组也是反对的，但不如第 2 组那么强烈，依次类推。特别要注意使每组之间的距离，看起来要相等。

③ 做出每个态度语的累积百分比分布表。

④ 做出每个态度语的累积百分比分布图。

根据累积百分比分布图找出中位数（与 50% 累积百分比相对应的横坐标的数值），该中位数为这一态度语的量表值。从累积百分比分布图上找出与 25% 累积百分比相对应的横坐标的数值，再找出与 75% 累积百分比相对应的横坐标的数值，两数之差（后者减去前者的数值），为鉴别态度语质量的主要指标。数值越小，说明测评人员对该态度语的判断越一致，即该态度语越确切；数值越大，说明态度语越含糊不清。当数值大于或等于 2 时，态度语就不便选用。

⑤ 选出量表值小且第 1、2 组内都有相应量表值的态度语 20~30 句，就构成了瑟斯顿态度量表，这是一个等距量表。

表 1-9 列出了一个用于测量妇女在经济界地位的态度量表。

表 1-9 妇女在经济界地位的态度量表

量表值	题号	态度语（项目）
2.5	1	妇女工作，使失业率增加
9.9	2	依靠男子的妇女，如同寄生物
7.0	3	在家庭外做事的妇女喜欢交往
7.6	4	如果家庭事情不多，妇女最好在外有事做
4.3	5	妇女的经济地位，不是一件紧要的社会问题
3.4	6	凡有幼年子女的妇女，不应在外做工
1.3	7	妇女的地位在家庭之中
9.3	8	男女在职业界上，应占同等地位
4.5	9	劳力的苦工，不得由妇女去做

续表

量表值	题号	态度语(项目)
5.1	10	妇女做事, 应只占次等的位置
10.6	11	政治上的重要位置, 应完全由妇女充任
8.8	12	应当鼓励妇女就业
2.8	13	工商业界聘用妇女, 会导致工作效率降低
3.8	14	做工的妇女, 失其娇态
0.1	15	任何工种都不应雇用女工
6.0	16	未婚的女子, 才准做工
10.4	17	如果工业都由妇女管理, 我们的一切经济问题, 皆可能解决
8.2	18	妇女有专职的, 其子女必羡慕
6.8	19	遇有必要时, 妇女应自食其力
0.6	20	妇女应该完全由男子供养
5.7	21	妇女是否做工, 应当视其家庭经济状况而定
6.4	22	妇女如情愿做工, 应准其做
7.8	23	未曾在外做工的妇女, 其眼光是很狭隘的
5.4	24	我们对妇女做工问题, 没有成见
1.6	25	妇女最重要的“职业”是生育子女

在实施测验时, 被试者就上述量表各项目表示赞成或不赞成。赞成的项目记作“√”, 不赞成的项目记作“×”。如果是团体测验, 不允许被试者之间相互交谈。填写答卷一般没有时间限制, 但为避免拖延, 普通时限为 15~20 分钟。计分方法实际上很简单, 一般将表示同意的项目依分数高低排列, 求出中位数, 以居中项目的量表值为该被试者的态度分数。如某一被试者对 2、4、8、12、17、18、23 诸句表示同意, 其相应的量表值, 如表 1-10 所示。

表 1-10 某被试者同样的项目之量表值

态度语	2	4	8	12	17	18	23
量表值	9.9	7.6	9.3	8.8	10.4	8.2	7.8

然后按上述量表值的大小排列等级, 求出中位数为 8.8, 如表 1-11 所示。

表 1-11 按量表值大小排列等级求中位数

量表值	9.9	7.6	9.3	8.8	10.4	8.2	7.8
等级	2	7	3	4	1	5	6

该量表规定分数越高越赞成妇女从事社会劳动，因此该被试者的分数 8.8 意味着比较赞成妇女到社会上工作。

瑟斯顿态度量表的信度一般为 0.8~0.9。

### (2) 李克特量表的编制方法。

李克特量表是评分加总式量表中最常用的一种，这种量表形式是李克特于 1932 年提出的。它是以肯定（正面）或否定（反面）的方式表述项目，然后假设每个项目之间都是等值的，均按反应中同意或不同意的程度统一赋分，所有项目得分的总和为个人态度的分数。

由此不难看出，李克特量表中所有的项目必须是同质的，即每个项目都是评定同一件事物（或同一个主题），编制的关键在于每个项目的选择是否妥当。具体的编制步骤如下。

① 针对态度对象选择量表项目，以肯定或否定方式表述。一般要求所选择的项目是相同性质的，因为态度间的相关性，所以可以利用看来与测评某一态度不是直接有关但实际相关的项目测评该态度。例如，人们对规范遵守的态度，与对规范合理性认识的态度是相关的，因此可以选用一些有关规范认识的态度语。

李克特量表的项目一般为 20~25 个。

② 试测。抽选一组有代表性的被试者，就上述初选出来的项目做出评价。通常分 5 个等级：“1”表示非常同意；“2”表示同意；“3”表示无意见；“4”表示不同意；“5”表示极不同意。

③ 赋分。以正面方式叙述的项目，越赞成给分越高。就 5 个等级类别的反应而言，非常同意的给“5”分，极不同意的给“1”分，中间的档次依次类推。以反面方式叙述的项目，越赞成给分越低。非常同意的给“1”分，极不同意的给“5”分。

④ 分析项目并进行筛选。首先，根据总分高低将所有被试者按顺序排列，将态度总分最高的 25% 被试者划为一组，总分最低的 25% 被试者划为一组；其次，计算此两组被试者在每项上平均得分的差异，即该项目的区分度或鉴别力。差异越大的项目区分度越高，越有保留价值。

如果有 5 个区分度都低且相等的项目，只需要从中删除 2 个，然后用统计学中的 t 检验对每个项目组的平均分与低分组的平均分之间的差异进行显著性检验，差异显著的项目保存，反之删除。

### (3) 哥特曼量表的编制方法。

哥特曼量表由单独且同质的少量项目构成。项目内容是对现象的描述,可以是短语,也可以是句子,数量不超过10个。每个项目有5个等级性的选项:①经常;②有时;③只有一次;④从来不;⑤不予以作答。要求被试者从中选择最适合他们情况的答案。当项目所描述的现象是正面的时,选①、②、③的得1分,选④、⑤的得0分。所有的项目得分相加为总分。

哥特曼量表的编制步骤如下。

① 设计出可用于揭示被试者特征的短语或语句。赞成的以“1”表示,不赞成的以“0”表示。

② 抽选一个有代表性的样组进行测试。假设某项目有15个人参与试验,其结果如表1-12所示。

表 1-12 哥特曼量表分析

被试者	测评项目								分数
	7	5	1	8	2	4	6	3	
15	1	1	1	1	1	1	1	1	8
7	1	1	1	1	1	1	1	0	7
9	1	1	1	1	1	1	1	0	7
10	1	1	1	1	1	1	0	0	6
1	1	1	1	1	1	1	0	1	7
13	1	1	1	1	1	1	0	0	6
3	1	1	1	1	1	0	0	0	5
2	1	1	1	1	0	0	0	0	4
6	1	1	1	1	0	0	0	0	4
8	1	1	1	0	0	1	0	0	4
14	1	1	1	1	0	0	0	0	4
5	1	1	1	0	0	0	0	0	3
4	1	1	0	0	0	0	0	0	2
11	0	0	0	0	0	0	0	0	0
12	1	0	0	0	0	0	0	0	1

③ 将被试者依总分从上至下排列,如表1-12中的“被试者”列与“分数”列。

④ 将项目按赞成次数最多至赞成次数最少从左至右排列,如表1-12中的“测评项目”列。

⑤ 去掉某些不易找出分割点的项目(指那些“0”),这样会给使用该测评软件的公

司带来误导，“1”不是集中于上下端，而是随机出现的项目），因为这类项目无法判断是被赞成还是被反对。

⑥ 计算交叉系数。首先，从表 1-12 中的分割线上找出“0”的个数；其次，从此表中的分割线之下找出“1”的个数。本例中，分割线上的“0”共 2 个，分割线下的“1”共 4 个，总计 6 个交叉数；最后，把交叉数代入公式  $C=1-(\text{交叉数}/0 \text{ 与 } 1 \text{ 的总个数})=1-(6/120)=0.95$ 。当  $C>0.09$  时，哥特曼量表成立。

从哥特曼量表分析被试者得分高低，可直接看出态度的差异情况，不会像前面两种量表那样出现分数相同而形态不同的现象。

---

## 1.3 人才测评应用中的认识误区

---

尽管现代人才测评的应用越来越受到各类组织的重视，但与此同时，许多人对人才测评还有各种各样的认识误区。

### 1.3.1 误区一：人才测评无用论

尽管现代人才测评的影响越来越大，但仍有人认为人才测评没用，甚至认为人才测评得不偿失，持这种观点的人总是有这样“充足”的理由：以往没有现代人才测评技术，企业不是也发展到今天了吗？但他们没有仔细想想，毕竟时代已经完全不同了，如果说科技人员的误用还可以弥补的话，那么管理人员的误用，特别是中高级管理人员的误用，将对组织造成无可挽回的损失。施乐公司曾对 500 名销售人员和经理进行测试，花费了 34 万美元，而实际增加的经济效益是 490 万美元；在 1998 年，美国电话电报公司投资 3.3 亿美元咨询费，重要工作之一是人才测评，综合效益是 15%，避免了 55 亿美元的经济损失。由此可见，科学的人才测评要比传统的选人用人办法准确、客观、有效。这就像医生给病人看病，过去只是看病人的脸和诊脉，这固然能在许多情况下正确地判断病情，但也经常出现判断失误的情况。而现在医生诊断病情时，会给病人化验血液、做 B 超检查等，这些先进的技术大大降低了病情判断失误的概率。现代人才测评技术就像化验血液、做 B 超检查等病情诊断方法一样，准确性要比传统的选人用人办法高很多。

### 1.3.2 误区二：盲目信任测评软件

从现有的企业人才测评的现状来看，人才测评多依赖于软件进行评估，测评人员对于计算机软件的使用过于频繁，导致其基本测评能力难以得到提升，测评工作名存实亡，人才的筛选也出现问题。盲目地追求测评指标，对于人才没有进行全面合理的评估。从

测评软件的发明来看,一些公司为了追求短期利润,没有经过合理的测试就将测评软件出售,不仅没有任何实践经验,而且存在较大的技术误区,超低的技术指标代表这样的测评工具误差较大、可信度较低,这样会给使用该测评软件的公司带来误导。

### 1.3.3 误区三:人才测评是选拔人才的最准确的依据

当前,人们认识到了人才测评是一种严谨、客观和准确的考查人的基本素质和能力的方法,有些企业就把人才测评看成人才选拔的最准确的或唯一的科学依据。这种看法虽然认识到了人才测评科学的一面,但过于片面,又产生了盲目的依赖。首先,人才测评的有效性和可靠性是具有一定限度的,不能单纯依靠测评分数准确定位一个人才,因此,测评必须和其他甄选手段结合使用;其次,人才测评主要考查人的知识和能力。但在实际工作中,决定人们绩效高低的因素不仅有知识和能力,还有激励水平,如果激励不足,测试评价很高的人仍然可能成为一个低效率的工作人员。因此,不仅要关注人才测评的结果,还要关注人才测评的应用,要全面、科学地看待人才测评。

### 1.3.4 误区四:对测评结果的准确性缺乏认识

许多组织对现代人才测评的测量结果缺乏正确的认识。有人认为人才测评就像血型分析那样神秘,也有人认为人才测评与物理测量一样准确可信,这两类看法都是不对的。一方面,我们必须认识到,现代人才测评通过对行为样本的测量来推测被试者某方面的素质,是建立在理论和统计技术基础上的科学评价方法,其测量结果通常要比传统的选人用人办法准确得多。另一方面,我们也必须承认,这种测量的准确性永远无法与物理测量相比。首先,人的测量要比物理测量复杂得多,在素质测评中对测量对象的界定往往不是十分明确的,如“应变能力”的测量,需要界定什么是“应变能力”,不同的人会有不同的看法,而不像体重、身高等物理测量对象那样有明确的、大家都认可的界定;其次,在素质测评过程中,经常会受到多种因素的干扰,特别是被试者自身因素的干扰,如被试者的紧张情绪、心情、身体状况等许多因素都会影响测评结果的准确性,这在物理测量中一般是不会发生的。由此可见,人才测评不可能是十全十美的,测评的结果也不可能100%正确。一般来说,能达到80%的准确性就令人满意了。

### 1.3.5 误区五:人才测评不如试用

有许多用人单位认为,再好的人才测评也不如试用的效果好,这个看法是不正确的。首先,试用在很多情况下是不现实的,我们不可能在一个职位上(如销售经理)同时试用多位候选人,以便考察他们谁更能胜任工作;其次,在某个职位上干得好不一定在另

一个职位上也干得好；最后，目前干得好，环境条件变化后未必还干得好。为了解决这些问题，我们将实际的工作场景进行浓缩，放在模拟的情景下，让被试者对各种各样的实际问题进行处理，在较短的时间内了解其素质，这就是情景模拟测试。大量的实践都已经证明，情景模拟测试具有较高的信度和效度，有很好的预测价值和经济价值。

### 1.3.6 误区六：测评一次，终身有效

有些企业进行人才测评是聘请的专业测评机构，一般的测评费用在 200~500 元/人，所以，企业为了节约成本，往往出现对某位员工进行素质测评后就决定其在公司的发展定位，“一次测评终身有效”。随着员工在工作中能力的不断提升，态度、价值观、兴趣、性格在不断变化，周围的环境在不断变化，所以，人才素质也在不断变化，人才测评应当具有动态性。当员工素质提高而现状得不到改变时，跳槽的可能性将增加。所以，“一次测评终身有效”的做法存在风险。企业应当在工作过程中，根据需要对员工进行动态测评，了解员工的素质变化，因材施教，员工也可以据此确定自己的发展方向。

---

## 1.4 人才测评应用中的问题与发展趋势

---

### 1.4.1 人才测评应用中的问题

当前我国人才测评事业的发展态势很好，但在应用中也存在许多问题，突出表现在以下几个方面。

#### 1. 人才测评技术的研究与创新严重滞后

现代人才测评在我国的应用源自心理测验，但是多年来，我们在心理测验的研究和自主开发方面做得不够深入，主要依赖对国外测验量表的修订，这导致目前我国人才测评机构中使用的大多数软件都只是简单的“舶来品”，几乎是从西方引进、修订的。然而，西方心理测验的文化背景未必完全适合中国人。因此，如何针对中国人特有的文化背景和心理特性，建立适合中国人的评价体系就显得至关重要了。

另外，对于包括面试在内的各种动态测评工具的创新也严重滞后，应付各种测评技术的培训力量非常强大，在这样的环境下，任何一种测评方法，如果不去创新都将面临失效的危险。以目前机关、企事业单位所广泛使用的结构化面试为例，广大被试者对这种测评手段已经非常熟悉，针对各类问题有很好的应答措施，如果不对结构化面试方法进行研究和创新，那么这种方法可能测量的不再是被试者的工作胜任力，而是“应试能力”，这是很危险的。



## 2. 人才评价服务水平低下

许多用人单位对测评技术重视不够,舍不得在人才招聘或测评技术方面的人力物力投入,随意委托没有资质和信誉的中介机构或个人从事人才测评工作,人才评价服务水平低下。就当前盛行的招聘考试来说,严峻的就业形势往往导致几十人,甚至几百人应聘一个职位,而一些单位对招聘考试的公平公正性缺乏认识,导致考试服务中错误不断。就最近的大规模招聘考试来说,有人随便从网上下载整套试题作为招聘笔试的试卷,导致数千人重考的考试事故;有的招聘考试中在几十分的主观题上附有答案,还解释说“由于疏忽忘了删除答案”,社会影响非常不好;还有的考试一张试卷上居然有10道重复题目……这些极其低级的错误让人感到非常痛心,它严重影响了人才测评工作的严肃性和客观公正性,影响了广大考生的切身利益。

## 3. 专业人才相对匮乏

人才测评涉及对人员素质的测量和评价,这对从事此项工作的人员有很高的责任意识和专业水平要求。在美国,必须由专业人士执行人才测评的操作,并对结果给予建设性的说明和解释。这种专业人员,在发达国家必须是博士,经过专业考核获得专业资格认证,还必须经过反复培训。而我国的现实情况是,专业人员相对于社会需求来说比较匮乏,于是一些对测评了解不多的人,去充当专家主持测评项目。长此以往,势必会损坏人才测评的声誉,阻碍人才测评事业的发展。其实,人才测评工作对从业人员的要求是很高的,人才测评工作者不仅要有较高的专业素养,而且要有高度的责任心、严谨细致的工作作风,甚至还要有吃苦耐劳的精神品质。当前,我们急需加强人才测评队伍的选拔和培养工作,扩充专业工作者数量,提高质量。

## 4. 人才测评服务行业秩序混乱

打开搜索网站,可以见到各种人才测评机构提供测评服务的广告,人才测评软件是支撑它们发展的重要支柱,但这些测评软件往往不是自主开发的。“营利”是许多市场化人才测评机构成立的根本目的。这种商业化运作的弊端是显而易见的。首先,会误人子弟。有的测评机构以节约为本,加上从业人员缺乏专业知识,引进了一些粗制滥造的测评软件,这些软件没有中国常模,问卷编制也很随意,被试者很容易被误导;其次,影响了行业整体声誉。在招聘会上,只需一台计算机,花几十元钱就能做一个测试,当场打印测验结果,告诉被试者适合什么工作,不适合什么工作。很多商家在利益驱动下对被试者很不负责任,自己不懂测评也敢对被试者指手画脚,这样势必严重破坏人才测评行业的良性发展。当前,急需有行业协会来规范人才测评服务的市场行为。

## 5. 相关法律法规缺位

有专家指出,导致人才测评发展受阻的根本原因是有关人才测评的法律法规不健全,管理跟不上。的确,在人才测评领域,至今尚无“行业标准”。一方面,任何一种测量工具,无须批准即可投入使用,其效果如何,却无人过问;另一方面,测量工具的优劣难以判断,导致在人才测评市场中,未经科学论证和测试,没有通过严格评审和认定的测评工具鱼目混珠。这样就造成测评结果失真,从而加深人们对测评的误解。因此,制定相关的法律法规,对人才测评市场进行有效的监督和管理成为当务之急。

首先,建立从业人员资格认证制度。从业人员的素质在一定程度上影响着该行业的发展,因此,规范人才测评市场,必须对从业人员进行资格认证。通过资格认证,按照国际惯例,参照国际标准,结合中国实际,一方面加强对从业人员的培训,提高现有从业人员的素质,另一方面吸纳更多优秀的符合认证资格的人员加入测评行业。

其次,完善中介机构管理。目前从事人才测评的机构还是以一些测评服务中介为主,因此,完善中介机构管理乃当务之急。其一,应该建立人才测评机构注册、审批制度,培育和发展人才测评市场。其二,对测评结果准确率低、弄虚作假,甚至违法乱纪的人才测评机构,要严加查处,情况严重的,取消经营资格,直至追究刑事责任。其三,对被试者的测评结果应予保密,保护被试者权益。

最后,加强技术专利保护。由于目前尚缺乏法规约束,各种盗版测评软件在市面盛行。因此,必须制定有关人才测评技术、专利保护的法规,保护知识产权,保障人才测评技术的发展。

### 1.4.2 人才测评的发展趋势

当前,人力测评的发展趋势呈现以下几个特点。

#### 1. 测评移动化

随着智能移动设备的普及,移动互联网时代已经到来。传统的人才测评行业当然也会受到它的影响。在移动互联网大潮中,对移动测评应用或服务的需求不断增强。

在人力资源顾问看来,为应聘者提供方便快捷的申请、应聘过程无疑是吸引年轻应聘者的重要手段。移动设备的灵活性与便利性也使得被试者几乎可以随时随地完成测评过程,比起传统的纸笔测试,甚至有监督的上机测试,组织测验的成本被大大降低了。

#### 2. 测评本土化

任何一个人才测评软件在中国的运用,都必须建立整套适合中国人的“常模”和“样

本”，这些不是从西方理论中照搬就能够做到的。可喜的是，在中国，一些本土化的人才测评工具正在兴起。比如，专注人力资源咨询和人才测评研究的倍智人才以个人能力和潜力作为底层逻辑，研发出新的中国化的人才测评工具。据介绍，“倍智人才”从个人先天潜能（如性格、驱动力和学习力等）和后天能力（如知识、技能、行为表现等）两方面对人才进行综合评定，并根据国人特有的生活背景来进行解释，最终与岗位进行匹配筛选，得出个人的胜任力。

### 3. 测评个性化

随着测评外包服务的实施，仅有测评软件是不够的，测评服务的概念越来越得到认同。购买测评服务的企业提出了更多的个性化需求，如素质模型的个性化定制、测评工具的个性化开发、测评实施的个性化设计、测评结果应用的个性化辅导等，这些都会被越来越多的企业提出来，而且标准会越来越高。

测评系统不能千篇一律，要根据企业的实际情况决定测什么、用什么测。在招聘新人时，不仅要考查应聘者的岗位匹配度，还要考查团队的匹配度，这一点是人才测评过程中容易忽略的问题。很多企业在招聘或者提拔员工时会考核其是否具有合作精神，但是没有考虑其与现有团队成员的匹配度。根据贝尔宾的团队角色理论，完美团队中的9种角色缺一不可，新吸纳进来的团队成员一定要角色补缺，所以，即使是招聘同一岗位，不同的企业需要的团队成员的角色风格不同，对其测评的偏重点也会有所区别。