

## 第十三章 智能的差异与测评理论

人的智能是多方面的，社会对智能的需求也是多样的，因而对聪明或能力的评判标准也应多样的。“智力”被认为是人类个体整体认知能力的反映。不同的人，其智力或能力是不同的，每个人都有自己的强项和弱项，它们构成了人类“个体智能”的五彩光谱，可让每个人去展现自己色彩斑斓的生命的彩虹。我们应该说，大多数人都是“聪明”的，只是他们“聪明”的领域和程度有所不同；特别是对正在发育中的儿童，更应如此看待。认识到每种“智慧类型”在未来的社会生活中都有可能发挥出其独特的作用，而且认识到每个人都有其各自的“智慧强项”，因此，我们对“智力”或“能力”进行测试和评估，主要目的应该是希望以此来发现各自的优势领域和优势潜能，并按照不同的智能光谱进行不同的个性化教育或做出不同的工作任务安排和分工。

一个人发展的水平并不完全由“智商”确定。“智商”也不是一成不变的，更不能决定人生的成就。教育的主要宗旨应该是为了启迪和开发人的各种潜能，因此，对智力的评估也应该只是一个获得个体智力发展现状和其各种潜在能力等信息的过程。其希望达到的目的，一是为其提供有益的反馈；二是为个体的成长提供合适的环境。我们研究差异的目的，主要是在正视人类个体智力存在差异的基础上，针对能力的差异采取不同的措施，或进行个别的教育指导，或充分发挥他们各自的特长，以促使他们的才能进一步发展。

对人类个体的智能发展进行指导和优化离不开对个体智能现有的发展状态和潜在的优势的认识。而要达到对个体智能的现有发展状态和潜在优势的认识就需要必要的测评。对个体智能的测评包括着对个体智力、知识、能力、以及其他生理要素和心理要素的测评。其中，最主要的是对个体智力、能力和心理素质的测评。本章，我们将研究个体智能差异产生的内在机制及主要表现形式；研究如何对个体的智能的现有的发展状态和潜在的发展作出评估，重点考虑的是进行科学测评的依据、原理和可行方法。

众所周知，智慧一直是哲学中的一个古老的话题，自柏拉图时代起，人们就一直着迷于对智慧的解释以及如何对它进行较好地测评。在心理学中，智力也有一个漫长的研究历史。自心理学成为独立的科学学科之日起，对智力和智能的研究就一刻也没有停止过，无数心理学家为此进行了不懈的努力。纵观整个智力研究的历史，可以看出，对智力本质的研究曾经历了三个取向的变化：20世纪60年代前，主要是采用因素分析的方法，试图找出智力的构成要素。其中有斯皮尔曼的智力二因素说、塞斯顿的智力多因素说、阜南的智力层次结构模式、卡特尔的流体和晶体智力理论以及吉尔福特的智力三维结构模式等；20世纪60年代至80年代，研究者吸取了信息加工的观点，试图为智力的运行机制构建出一个合理的模型。其中典型的成果有斯腾伯格的智力三元理论，它较好地揭示了智力的心理机制；80年代以后，人们对智力开始进行多层面的研究，试图使智力更加人文化、生活化。其中具有代表性的成果有加德纳的多元智能理论和情绪智能理论等。这些关于智力的探讨

虽说对智力的本质都有所触及，但在实际应用过程中却都发现了不可避免的局限性。这些表面上看来很完美的理论实际上都存在一个问题：智力到底是如何产生的？！显然，这些思辩性较强的理论很难作出令人满意的回答。

人们对智力和能力的兴趣不仅仅在于对其进行理论上的探讨，更重要地是希望对它们进行有效地测量和评估并有针对性地去指导个体的发展。心理学中对智力的测量研究已有近百年的历史了。然而，在这方面所取得的成绩却难尽如人意。即使是现代智力测验的两个重要代表人物——比纳和韦克斯勒，也都承认已有智力测验的不足。波林曾提出，智力就是 IQ 测验所要测量的东西。表面上看来，这是智力的操作定义，而实际上，这也是一种理论上的无奈。不少智力研究者对智力的测评至今仍显得十分迷茫，究其原因，我们认为，对于原本复杂的智力现象总是抱着过分简单化的观点应是一个很重要的方面。

过分简单化的智力观所导致的是智力测评的狭隘，是智力测评与智力理论的严重脱节。很大程度上，已有的智力测评很多是在理论的“沙滩”上发展的。我们不能说它没有一点理论基础，但是，它们确实没有坚实的理论基础。它们所测得的，只是人类智能的一部分，只是测得了人类的一些分析性能力等。这就是为什么现有的智力测验根本无法对一个人的发展做出准确的预测，也就无怪乎人们对于智力测验质疑的声音会越来越大。

智能理论是关于智能的本质和产生的内在机制的理论，智能的测评理论则是对智能的个体差异进行测评的基础理论。它要研究人类个体智能差异产生的原因，研究对个体智能的差异进行测评的基本理论和方法，包括关于个体智力的测评理论和方法、关于个体知识与能力的测评理论和方法以及关于个体心理素质的测评理论和方法等。其中，个体智力测评包括对个体先天“遗传基质”的测评（通过测评以确定与个体智力有关的先天素质究竟如何？）和对个体智力发育状况的测评（通过测评以确定个体智力发育状况是否正常？）；关于个体知识的测评要通过测评以确定个体已经知道了什么或掌握了什么程度；关于个体的能力测评要通过测评以确定个体能干什么或干的水平如何；关于个体的心理素质测评要通过测评以确定个体的相关心理素质如何。

智能是对主体（主要是人类个体）信息处理能力的一种表征，是主体的一种可进行信息加工的综合能力。由于主体具有的用于信息处理的各种基本素质、知识水平和实际操作能力各不相同（它们要由先天的基质与后天的发展训练决定），要处理的信息的类别各不相同，处理不同的信息所需要的素质、能力和知识各不相同，因此，要统一评价所有主体的智力水平显然是很困难的，也是没必要的。我们反对用一个统一的“智商”来评估“所有的人”，因为人才不可能“千人一面”。古人尚且希望天公“不拘一格降人才”，主张“英雄不问出处”，何况“日新月异”的现代。我们也并不认为智力和能力的测评毫无用处，因为“育人”需要“懂人”，“知人”才能“善用”，关键只在于如何准确测评。按说，主体的知识、智力和能力水平，应有相对水平和绝对水平；但是，要我们对主体的各类能力给出一个统一的测度和评价，是不现实的。我们只能针对不同的测评目的和特定的信息处理要求，对主体的特定的（智力或能力）水平做出一个有针对性的大概的评估。这类评估只能是一类模糊评估。另外，由于评估时的“测试数据”与实际工作时的具体问题处理要求是不一样的，故评估出的水平不一定就是实际工作时的水平。评估只是一种“参考测度”。

### 13.1 关于人类智能的个体差异及其产生的内在机制研究

#### 13.1.1 智能的差异是客观存在

客观地讲，智能的差异是一种客观存在。不同主体有不同的智能。也正是因为存在着种种的差异，才使得我们要研究智能。智能的差异表现在多个方面：一是不同生物种群之间存在着明显的智能差异；二是人类社会中不同人群之间，比如男女之间、儿童与成人之间，存在着明显的智能差异；三是同类人群的不同个体之间存在着明显的智能差异；四是不同个体的不同智能构成组分之间也存在着明显的差异。下面，我们仅列举出几类比较典型的存在智能差异的例子。

• **不同生物种群之间智能的差异。**有人认为，只有人才具有智能，动物没有智能；也有人认为，人和动物都具有智能。我们比较倾向于后一种观点。但是，不管是从那一种观点来看，人和动物之间在智能方面存在着差异却是非常明显的，或者说不同生物种群之间在智能方面的差异是非常明显的。尽管人们有时也会“夸奖”某某动物如何如何“聪明”，但很少有人真的会认为动物能“聪明”到超过了人的程度。尽管我们赞成说“动物也有智能”，但是，这里所说的“智能”是应该有等级、有层次的。智能等级的高低，首先应该取决于其产生的基质，也就是不同生物种群的脑神经系统的发育；也取决于其可处理问题的难易程度或处理问题时的“聪明”程度。为此，有人曾建议引入了一个“智能度”的概念来刻画（不同主体）智能的等级，由此，我们可以认为，人类的智能具有最高等级的智能度，灵长类动物的智能具有相对较低等级的智能度，而其它生物的智能将具有更低层次智能度。

不同生物种群之间在智能方面的差异，主要是由于它们具有不同的“思维”器官，也就是说，主要是由于不同生物种群的脑神经系统不同而引起的。地球上的不同的生物种群，是自然界长期进化的结果，其不同的“思维”器官，当然也是自然界长期遗传和进化的结果。它的基质，主宰着不同生物种群智能发育的极限，也从根本上决定了不同生物种群智能的高低。

• **不同人群之间智能的差异。**不同人群之间是否存在差异？如果公正客观地看待智能差异问题，我们相信很少有人会怀疑不同群体之间智能存在差异的结论。一个明显的例子是儿童与成人之间在智能方面所存在的明显的差异。还有人认为，男女之间也存在着智能或者说智力方面的差异。

是男孩聪明，还是女孩聪明？这历来是个有争议的问题。不少人认为男性比女性聪明，他们的论据是脑力劳动者男性居多，世界上有成就的人大多也是男性。也有人认为女孩聪明，因为从幼儿园到小学，女孩在学习上进步快，表达能力强，成绩也要比男孩好。但不管怎样认为，有一点是大家公认的，那就是：男女之间在智力或智能结构方面存在着差异。对此，科学家曾作了大量的试验、分析和研究。比如，上世纪20年代，美国心理学家桑戴克曾通过实验得出结论：男性和女性的智力结构不同，各有自己的智力优势。对美国高中学生的学习成绩的调查表明，男生在数学和物理学方面要超过女生，而女生则在语文和外语方面遥遥领先。日本的心理学家也做过调查，结果显示：在方向和位置的辨识、图形的组合等方面，男性优于女性；而在语言表达能力、会话的流畅性、记忆力和处理人际关系等方面，女性要比男性强。

关于智力的性别差异，目前研究较多，而且结论各异，但基本一致的结论有两方面：（1）男女智力的总体水平大致相等，但男性智力分布的离散程度比女性大，即很聪明的男性和很笨的男性都比女性多，智力中等的女性比男性多；（2）男女的智力结构存在差异，各自具有自己的优势领域。男性的视知觉能力较强，尤其是空间知觉能力，男性明显优于女性。女性的听觉能力较强，特别是对声音的辨别和定位，女性明显优于男性。男性偏于抽象思维，喜欢数学、物理和化学等学科。女性长于形象思维，喜欢语言、历史、人文地理等学科。一般地说，女性比男性口语发展早，在语言流畅性及读、写、拼等方面均占优势，但男性在语言理解、言语推理等方面又比女性强。麦科比

(E. E. Maccoby) 研究发现, 对小学生, 在语文能力方面, 女生一般优于男生; 在数学能力方面, 男女生无明显差异; 在空间关系(方位与方向)判断方面, 男生一般优于女生。根据麦科比的研究, 以及此后其他教育心理学家们的验证, 一般认为, 小学阶段之后, 女生在语文方面的优势逐渐消失, 而男生在数学能力方面的优势却继续增加。这种学业性向的性别差异转变的现象, 到中学以上直至大学阶段会更趋明显。

智力的团体差异可能是最有趣而又最敏感的问题。已有研究表明, 男女两性的总体 IQ 水平没有显著差异, 但在特定能力上存在差异, 男性在视空间和数学能力上有优势, 女性在言语能力上有优势。有人认为, 荷尔蒙水平以及社会化要求的不同, 是造成智力性别差异的主要原因。智力的种族差异有着复杂的模式, 很难综合概括。以美国人为例, 有研究认为, 美国黑人的 IQ 明显低于白人, 拉丁裔美国人的 IQ 介于白人和黑人之间, 中、日等亚裔美国人的 IQ 虽不比白人高, 但他们的学业成绩、职业成就总体水平较高。但是, 还没有直接的证据表明是遗传造成了智力的种族差异, 有人认为, 可能是经济地位、社会等级、不同群体的亚文化等因素才导致了这种差异。

那么, 男女智能的结构性差异是如何造成的呢? 一些心理学家认为, 这是因为男女的感知觉能力不同的缘故。女性的嗅觉、触觉比较敏感, 对声音的辨别和定位也不错; 相比之下, 男性在视觉辨别方位方面则稍强一些。他们在研究追忆、认知与信息输入通道(视觉通道、听觉通道、触觉通道等)关系时也发现, 无论信息从哪一通道输入, 女性在追忆和认知两种记忆类型上都比男性好。也有专家认为, 男女智能的结构性差异的根源在于社会和大脑的功能。社会对男性和女性的培养和期望是不同的, 他们从幼年起就被要求学习和选取不同的活动类型。例如, 在一些家庭中, 男孩被带到室外参加活动的机会要多一些, 他们也就喜欢做跌、爬、滚、摸等活动量大的游戏, 女孩则较多地留在家里玩耍, 再加上孩子多倾向于模仿同性成人的行为, 于是, 通过这些不同类型的活动, 男孩和女孩的能力就出现了差异。

科学研究人员通过各种试验还发现, 人的大脑功能的确存在着性别差异。这也许是男女智能结构存在差异的一个重要原因。他们的研究成果揭示, 在男性大脑内, 与语言功能相关的颞叶脑平面总是左脑明显大于右脑, 可见男性的语言中枢是集中在左脑的。而在女性大脑内, 这种差异远不如男性那样明显。看来, 女性的语言功能并没有集中在左脑。因此, 他们认为, 女性在处理语言信息的时候, 很可能是利用了联系左、右脑的神经线路, 是由左、右脑互相协调来完成的。女性大脑的这种通道总是多于男性, 女性喜爱“喋喋不休”的原因也许就在于此。

也有人提出“与众不同”的观点。他们认为, 归根结底, 是性激素造成了男女之间不同的智能结构。在成长过程中, 男性会分泌大量的雄性激素, 这些激素会延缓左脑的发育, 使右脑发达。女性大脑的早熟则与雌性激素有关。由于脑早熟在女性身上出现得早, 因此与控制语言有关的皮层也就更早发生作用, 女性就表现出了语言方面的优势。

男女智能结构的差异, 究竟是由于感知觉能力的不同, 还是因为社会的培养与期望的不同以及大脑功能的差异, 或者是由于性激素的作用而造成的? 目前还缺乏统一的认识。也许这是多种因素共同作用的结果, 但其最主要的因素是什么呢? 它们是怎么协同作用的呢? 对此, 神经生理学家和心理学家至今还无法明确回答这些问题。在这里, 我们并不想更深入地讨论这些问题。我们只是把它作为一个事例提出, 说明不同人群之间确实存在智能差异。

• **人类不同个体之间智能的差异。**人类不同个体之间存在着智能方面的差异也是一个不争的事实。这也是我们在考察和研究智能的差异时要重点考虑的问题。人类不同个体之间在智能方面的差

异主要表现包括：智力方面的差异、知识方面的差异、心理素质方面的差异和综合能力方面的差异等。

当我们在从事一些具体的活动时，由于这些活动的内容、性质和目的要求不尽相同，也由于各人对活动的内容、性质和目的要求的认识 and 理解的差异，出现不同的活动结果应该是很正常的事。假若我们所从事的是相同的活动，并且这一活动的内容、性质和目的要求都已十分明确，并假定从事这一活动的所有的人对此活动的内容、性质和目的要求的认识和理解也完全一致，但由于各自采用的方法、手段和途径等可能不尽一致或不尽相同，也会出现不同的活动结果，或达到目的的时间各不相同，这就反映出了各个个体从事此项活动时的智能存在差异，或者说有高低之分。

比如，同一教研室中两位资历相同的教师，给学生讲授相同课程的相同章节的内容之后，得到的评价会有总体的和具体方面的好、一般或较差；又如，同一个年级的学生同时解答同一道数学题，有的学生很快就算出结果，有的则需要一定时间方可解答出，有的学生却几经求解也不能完成这道题的计算。应该说，这些学生都有解答此数学题的基本知识，而他们完成此题的时间有先有后，就反映出他们在解决这一问题上的能力方面存在差异。即便是我们只让几个优生来同时解答同一道数学题，他们的思维方式、解题方法也不一定是一致的，完成这一任务的时间仍可能会有先后之分，完成任务的质量也会有差异。这就说明，在我们的各项实践活动中，个人与个人之间，能力总是有高低之分的，这是客观的事实。

从中国古代的因材施教思想到近代西方的建构主义思潮，无论是心理学界还是教育学等领域，研究者们无不在强调了解个体差异的重要性。造成个体差异的一个最为重要的因素即在于个体的智力水平。自从 19 世纪末 20 世纪初高尔顿和比奈等人开创当代智力心理研究以来，对智力的探索已取得了许多进展。无疑，这在很大程度上促进了我们对个体差异的更好理解。

不过，需要说明的是，作为一个正常的人，我们人人都具有从事高等智能活动的的能力，只不过在不同种类的活动中表现各不相同罢了。对于此项活动，你可能胜人一筹，但对另一项活动，他又可能高你一码。不同个人能力方面的差异，在很多时候是由于各人的知识和智能结构不尽相同而造成的。

智能的差异有时也表现在专业性活动方面。在各类专业性活动中，一些人的智力和能力的要素可能会优先得到充分的发挥从而表现出特殊才能，这种会表现出特殊才能的个体常被称为高智能“人才”。比如，优秀的企业家是高智能人才，杰出的音乐家、舞蹈家和科学家也是高智能型人才。这些高智能型人才，一般更具有创造能力和适应能力，在专业性工作中往往超过一般人的能力。

### 13.1.2 智力常态人群个体之间智能差异的主要表现

人类个体在智能方面存在多种差异，这些差异可以从多个方面加以分析。人类不同个体之间智能方面的差异包括智力方面的差异、个人知识方面的差异、心理素质方面的差异和综合能力方面的差异等。在考虑个体智能差异时，人们最关心的是：

(1) **人类个体智力或能力发展水平的差异。**研究发现，就一般而言，在所有人中，智力发展水平都基本呈常态分布，即智力极低或智力极高的人都很少，绝大多数的人的智力水平都属于中等智力水平。

有人根据人类个体智力发展水平的差异，将儿童分成三个等级，即超常儿童、常态儿童和低常儿童。超常儿童是指其智力发展或某种才能素质显著超过同龄儿童平均水平的儿童。他们的“智商”一般在 130 分以上；其共同的心理特征表现为：浓厚的认识兴趣，旺盛的求知欲；思维敏捷，理解

力强，有独创性；敏锐的感知觉，良好的观察力；注意力集中并易于实现转移，记忆速度快而准；进取心强，勤奋，有坚持性等等。低常儿童是指其智力发展水平明显低于同龄儿童平均水平并有适应性行为障碍的儿童，又称智力落后儿童。他们的“智商”一般在70以下。低常儿童的主要特征为：知觉速度缓慢，范围狭窄，记忆能力差，语言发展迟缓，词汇贫乏，思维概括能力差，生活自理能力差等等。不过，需要指出的是，智力发展水平的差异并不完全由“智商”确定，因为造成个体智力发展差异的原因很复杂，有先天因素，有后天因素，也有认识因素。我们不赞成单纯以“智商”来作为评定智力发展水平的标准。我们研究差异的目的，也主要是在正视个体智力存在差异的基础上，针对个体能力的差异采取不同的引导措施，“因材施教”，通过个别化的教育来充分发挥他们各自的特长，以促使他们的才能得到进一步发展。

**(2) 人类个体智能类型方面的差异。**个体智能在类型方面的差异是指构成个体智能的各种能力和素质方面存在的差异，主要表现在知觉、记忆、想象、思维等认知能力方面的差异和心理素质方面的差异，以及知识掌握方面的差异等方面。有人认为，个体知觉主要有三种类型：综合型，即知觉具有概括性和整体性，但分析能力较弱；分析型，即知觉具有强的分析能力，对细节感知清晰，但整体性较差；分析综合型，具有上述两种类型的特点，即同时具有较强的分析能力和概括能力。个体的记忆能力也有不同类型，根据记忆材料方式的不同，可分为：视觉型，运用视觉记忆效果好；听觉型，运用听觉识记效果好；运动型，有运动参与时记忆效果较好；混合型记忆，运用多种记忆方法时效果较好。在言语和思维方面，个体也存在类型差异，有的人言语富于形象性，情绪因素占优势，属于生动的言语类型或形象思维类型；有的人言语富于概括性，逻辑因素占优势，属于逻辑联系的言语类型或抽象思维类型；还有居二者之间的混合型。在思维能力方面，每个人在思维的深刻性、灵活性和批判性等品质上又都有自己的特点。能力的类型差异，不一定就代表其智力水平的高低，但会影响个体学习的过程和获取知识的方式。

**(3) 人类个体才能显现早晚的差异。**个体能力不仅在“质”或“量”的方面可表现出明显的差异，其能力表现的早晚也存在着明显的差异。其能力的显现可有“早慧”或“大器晚成”的差别。

很小就显示出不凡的能力称为早慧。能力早期显露的事例，国内外不乏其例。许多名人在幼年时期就已显露了其不凡的才华。李白“五岁诵六甲，十岁观百家”；莫扎特五岁开始作曲，十二岁创编歌剧；控制论的创始人维纳，四岁能自由地阅读书籍，十一岁写出论文，十四岁大学毕业，十八岁就获哈佛大学博士学位。有研究表明，能力早期表现在音乐与绘画领域中最为常见。据哈克(Haecker)、齐汉(Ziehen)的统计，儿童在三岁左右开始显露音乐能力的情况最为普遍。

没有在早期就显露出才华并不意味着将来不可能大有作为。与能力早露者相对应，“大器晚成”者也不乏其人。事实上，古今中外，“大器晚成”者也大有人在：姜子牙辅佐周武王，72岁才任宰相；著名画家齐白石，40岁才表现出绘画才能；人类学家摩尔根发表基因遗传理论时年已60岁。造成“大器晚成”的原因很多，有些可能是因所专攻的学术领域具有某种长期性，不可能一蹴而就，需长期努力；有的可能是因为个人早期不够努力，后期才加倍勤奋的结果；有的可能是小时智力平常，通过长期的主观努力，智力像菊花一样，到了人生的秋天才显示出绚丽的光彩；也有些可能是某种特殊社会原因或机遇造成的，早期由于机遇和环境的限制，缺乏相应的发展条件，直到晚年才有了施展才能的机会。

一般来说，科学家作出最大贡献的最佳年龄是中年。有专家认为，中年人年富力强，精力充沛，既有丰富知识经验，又有较强的抽象思维能力和记忆能力，思维敏捷，较少保守，勇于创新，是出

成就的最好时机。有人对 301 位诺贝尔奖获得者做了统计, 结果表明, 30~45 岁是人的能力的最佳年龄区, 诺贝尔奖获得者中有 75% 的人的创新年龄处于这个最佳年龄区。这是因为重大的发明创造除了需要良好的智力品质和充沛的精力以外, 还要有丰富的基础知识和实际经验以及符合该领域要求的高水平的专门能力和创造能力。这就需要一段时间来积累知识、经验和发展相应的能力。中年人既有较强的抽象思维能力和记忆能力, 又有丰富的基础知识和实践经验, 所以, 中年期是个人成就最多、对社会贡献最大的时期。研究还表明, 从事不同学科的人, 其最佳的创新年龄阶段也有不同。创新有个最佳年龄阶段, 但并不是说人在这个年龄阶段之外就不可能有所创造、有所发明。有人“早熟”, 就有人“大器晚成”。另外, 随着社会的进步、科学的发展和教育质量的提高, 创造的最佳年龄有向两端延伸的倾向。

需要说明的是, 人的各种不同的能力在发展速度上也是不同的。某些能力发展得较早, 有的却较晚; 到了老年, 各种能力的衰退速度也是不一样的。有研究表明, 知觉能力发展较早, 也最先开始下降; 其次是记忆力, 然后是思维能力。比较、判断能力 80 岁后急速下降, 动作反应速度在 18~29 岁时可发展到最高峰, 在以后一段时间内仍可保持较高的水平。从表 13.1.1 中可以看到, 在 18~49 岁这个年龄阶段, 四种能力的发展水平几乎都处于最高水平, 尤其是比较和判断能力的发展水平是最高的。

年龄	10~17岁	18~29岁	30~49岁	50~69岁	70~89岁
知觉	100	95	93	79	46
记忆	95	100	92	83	55
比较、判断	72	100	100	87	69
动作反应速度	88	100	97	92	71

表 13.1.1 不同能力在不同年龄阶段的不同发展水平

(4) **人类个体能力组合方面的差异。**个体能力组合方面的差异是指为完成某一活动所需要的各种能力的不同结合方式的差异。个人在不同能力上的差异是很明显的, 从而构成其特定能力的各种组合因素也是不一样的。为了顺利完成某一项活动, 可能需要由多种能力的各式各样的结合来实现。各种不同能力的综合, 也就形成了人与人之间能力上的“千姿百态”的差异。

### 13.1.3 关于人类智能个体差异内在机制的理论研究<sup>[1301]</sup>

对个体智能差异研究最多的是对个体智力方面的差异的研究。众所周知, 人类个体存在着明显的差异, 智力差异就是其中的一个重要方面。智力个体差异研究的历史几乎与心理学的历史一样长久。在近百年的研究过程中, 研究者从不同的认知出发, 从各种角度对各种人群进行了广泛的研究, 试图对智力个体差异的内在机制做出更深刻的解释, 从而也形成了多种不同的理论, 如心理要素差异理论、参数差异理论、策略取向和适应性理论等。它们都试图从不同角度对智力的个体差异做出解释。

目前, 人们对人类个体智力及其差异的研究包括外显理论和内隐理论。其中, 外显理论是指研究者根据对研究对象的直接研究所提出来的, 通过实验性技术形成和发展起来的(有数据的验证和支持), 并在出版物和专业会议中进行传递和分享的理论。内隐理论指人们在日常生活和工作背景下形成的, 并以某种形式保留在人们头脑当中的关于某些事物的看法。智力的内隐理论主要通过调查来总结出一般人心目中“高智力者”的特征, 如, 解决问题的能力、言语能力、社会实践能力等。智力的外显理论有许多形式, 其中最具有影响的有两类理论: 差异理论和认知理论。智力的差异理论以研究人类的个体差异为基础, 根据一系列主要能力(如语言能力、推理能力等)来理解智力。差异理论主要以静态的智力结构分析一因素为基础。而智力的认知理论则认为智力主要是人的认知能

力的一种综合表征。

关于智商常态人群智力个体差异的认知,目前还存在各种不同观点。有一般能力理论,认为是G因素的不同决定了特殊能力的分化程度和差异;有认知机能差异理论,认为是认知机能系统中的一个或者几个认知机能缺陷造成了智力功能的差异;有认知结构差异理论,认为是个体认知发展结构中的要素功能的不同和要素关联关系的强弱影响了智力的差异;还有文化活动结构差异理论、基于遗传进化的神经效率差异理论、个体神经活动特性水平和神经活动区域效率差异理论、人格中介情景任务和人格限制成分功能表现差异理论,等等。自智力测量出现之后,智力个体差异研究逐渐进入了实证时期,也形成了很多关于实证性智力个体差异的认识。

### 13.1.3.1 关于智力差异的心理要素研究

智力差异的心理要素研究指的是以心理测量学理论为基础,运用因素分析的方法对智力个体差异进行探讨的研究取向,该类研究又被称为心理测量理论。其基本观点是,个体的智力由“潜在”的心理因素决定,并且可以从一些特定测验的成绩中体现出来。这一理论中各派子理论的不同点在于所强调的因素数目和这些因素组织的形式等。

智力是人顺利完成某种活动所必需的并直接影响活动效率的个性心理特征。一般认为,人的心理因素可分为两类:一类是与认知过程,如感知、记忆、思维、想象等等心理活动直接有关的,称为智力心理因素或认知心理因素。它们主要是个体在认识客观世界的过程中形成的一系列稳定的心理特征,包括观察能力、注意能力、记忆能力、思维能力和想象能力等,是直接参与认知过程的基本心理要素,是最能体现人的智慧水平的心理要素。智力心理要素在智力心理活动中直接参与了对客观事物的认识,是智能行为的基础,属于活动中的操作-执行系统。智力心理活动要素在学习活动中对各种知识起着吸收、加工和处理的作用。另一类是与认知过程的启动、维持、调节有关的,如兴趣、动机、情感、意志、性格、态度等,称为非智力心理要素或非认知心理要素。非认知心理要素是个性心理要素结构中相对于认识心理要素而言的。关于非认知心理要素的组成,心理学家的意见并不完全一致,大致包括兴趣、情感、意志、性格、道德、思想和态度等七方面的心理活动,它们在人的行为的或心理的活动中起着动力性和方向性的作用,属于活动中的动力-调节系统。也可认为,非认知心理要素主要指人的个性结构(动机、态度、兴趣、性格、气质)、情感结构(情绪、情感)、道德结构(诚实、正义、友爱、勇气)和意志结构(独立性、勇敢性、坚持性、自制力、自我平衡力)等。

对于智力,我们可以从各个方面进行解读。若从理性哲学观点出发,可认为智力主要是指抽象思维能力;若从教育学的观点出发,可认为智力主要是学习能力;若从生物学观点出发,可认为智力主要是适应新环境的能力。中国较多的心理学家认为,智力是指认知方面的各种能力,即观察力、记忆力、思维能力、想象能力的综合,其核心成分是抽象思维能力。若从智力发展的角度考虑,则智力可有两种涵义:其一是指已表现出来的实际能力或已达到的某种熟练程度,可用实际测验来测量;其二是指潜在能力,即尚未表现出来的心理能量,是通过学习或训练后可能发展起来的能力或应该能达到的某种熟练程度,可用意向测验来测量。心理潜能是一个抽象的概念,它是各种能力展现的可能性,只有在遗传与成熟的基础上,通过学习才可能变成实际能力。但是,一个人的实际能力和心理潜能是一个不可分割的统一体。心理潜能是实际能力形成的基础和条件,而实际能力是心理潜能的展现。

人的能力存在个体差异,有一般能力方面的个体差异,有特殊能力方面的个体差异,有能力成

熟早晚的个体差异,也有在不同环境下的不同表现方面的差异等。从原则上讲,这些差异可通过一个人实际工作或学习的成就来进行考查,也可用各种形式的能力测验或智力测验结果来揭示。影响个体智力和能力形成与发展的因素也很多,概括地讲,可有三方面:一是先天性的因素,包括遗传基因和胎儿期因母体体内环境影响所造成的个体“先天”特征;二是后天环境因素,包括家庭影响、学校教育和社会历史文化的作用等;三是个人的主观努力。在能力形成的条件问题上,大多数人都接受遗传与环境两种因素共同决定的观点,认为遗传和环境在能力的形成和发展中都是必要的条件,能力决定于遗传和环境的共同作用。但我们认为,还应考虑个体的主观努力。

自20世纪20年代以来,心理学家们致力于研究智力的构成要素,先后形成许多智力结构的理论,其中有英国心理学家C.斯皮尔曼的二因素论、美国心理学家E.L.桑代克的多因素论、美国心理学家L.L.瑟斯顿的群因素论、英国心理学家P.E.阜南的层次结构论、美国心理汉弗莱斯的智力二维结构模型和美国心理学家J.P.吉尔福德智力三维结构模型等。其中,斯皮尔曼的一般因素智力理论以G+S智力结构理论为基础,认为影响智力功能差异的因素有两个方面:一是主要受遗传影响的一般能力的高低差异;二是主要受实践影响的个别活动领域的特殊能力优劣的差异。一般能力潜在地决定了特殊能力可能达到的水平,实践环境则决定了特殊能力现实的水平,智力水平低的群体能力结构分化程度也低,特殊能力之间的独立程度随着一般能力的功能提高而提高,一般能力功能高低对在同环境下的个体间的智力水平个体差异、能力结构分化状态具有潜在的决定作用。在相同的一般能力功能水平上,环境差异则现实地决定了特殊能力的水平和能力结构分化状态。认为,智力的个体差异反映了G因素功能的高低以及受一般能力因素制约的不同群体智力结构的特殊能力的分化程度的不同。多因素结构理论则否认一般能力的存在,认为智力差异反映了各种基本心理能力倾向差异或功能搭配的差异。Thurstone和他的追随者们通过研究提出了群因素论,主张用7个因素可以解释个体的智力差异,即言语理解、词语流畅、算术能力、空间关系、记忆广度及保持时间、知觉速度和归纳推理。至今,许多研究者依然在利用这些因素或者它们的派生物来说明个体差异。Thurstone认为,智力个体差异属于不同水平的各个基本心理要素的搭配差异,个体的智力结构与个体对活动的选择和兴趣具有密切的联系,独特的智力结构具有了不同的功能水平。J.P.吉尔福特(Guilford)提出的智力三维结构模式则认为,智力结构内不存在普遍限制作用的一般能力,而只有由内容维、操作维、产物维构成的多种基本能力;在基本能力结构下,个体间的基本能力差异造成了每个个体都具有独特的基本能力结构,独特的基本能力结构也就决定了个体的智力功能差异。智力是由多达120多个的因素组成的,这120多个因素可按照三个维度来分类,即操作(operations)、内容(contents)和产物(products)。操作,即思维的方法,可分为认知(发现或认识)、记忆(保持)、发散性思维(或称求异思维,以不同的思维方法求得新的答案)、集中性思维(或称求同思维,指向一定问题而得出一个正确的或最好的答案)和评价(判定知识的适合性)五个项目。内容,即思维的对象,可分为图形(通过感官看到的具体信息)、符号(单词或数字)、语义(言语含义或概念)和行为(与人交往的智力)四个项目。产物,即把某种操作应用于某种内容的结果,可分为单元(单词、数字或概念)、类别(一系列有关的单元)、关系(单元与类别之间的关系)、系统(用逻辑方法组成的概念)、转换(改变,包括对安排、组织和意义的修改)和蕴含(从已知的信息中观察某些结果)六个项目。三个维度的任何一个项目之间的相互结合,即可得出构成智力基础的 $5 \times 4 \times 6 = 120$ 个因素。英国心理学家H.J.艾森克也提出了关于智力的一个三结构模式。该模式包括三个维度:心理过程、测验材料和能量。艾森克的“心理过程”概念与吉尔福特的智力三维结构模式中的“操作”概念相似,艾森

克的“测验材料”与吉尔福德的“内容”相似。艾森克的“能量”则包括速度和质量，速度指被试在智力测验中的反应速度，质量指改正错误的多寡与解决问题时的坚持性等。有人认为，艾森克的“能量”也许更能反映被试成绩的高下。

20世纪70年代以后，研究者在运用多变量及因素分析技术方面越来越成熟，开始进入内部修改和元分析时期。比如，Dunnette在总结相关研究的基础上指出，其中有10个因素：中止速度、流畅性、归纳推理、联想记忆、记忆广度、数字能力、知觉速度、演绎推理、空间定向和言语理解，已经经受住了时间的考验，认为它们代表了相对稳定的认知心理因素。以晶态和液态智力结构为基础，Cattell和Horn等指出，智力功能个体差异的来源有可能是受遗传影响的“液态智力”、受文化环境影响的“晶态智力”以及三类“速度因子”的差异等三个方面。1993年，Carroll对10年内所发表的研究数据进行了再分析(re-analysis)，这种再分析是对分析数据的分析，类似于现在的元分析(meta-analysis)，其中涉及到了2500组数据、覆盖了10000个变量。在这个过程中，Carroll运用阶层探索性因素分析(hierarchical exploratory factor analysis)的方法来完成他的工作，通过对诸多组群因素的再分析，他把智力划分成了一个三水平的层级结构。最上面一级水平是一般智力G；中层是8个能力范畴：晶体智力、流体智力、一般记忆和学习、视知觉、听知觉、提取能力、认知速度、加工速度；这些一般能力的每一种类型又被进一步细分为各种特殊的初级水平因素，例如记忆被分为联想记忆、情节记忆、记忆广度和视觉记忆。尽管Carroll自己也承认，还可以再增加一些其他因素，特别是在初级水平上；而且，不同水平、因素之间的联系也可能比他划分的还要复杂。但是，从目前来看，这可能已经是争议最少的一种划分了，它是当今因素取向思想最有代表性的总结，已经得到了不少人的赞赏。

### 13.1.3.2 关于人类个体智力差异的认知加工参数研究

随着实验心理学和认知心理学的发展，心理学家一度对认知加工的一般过程更感兴趣，忽视了个体差异研究。20世纪70年代，亨特等人通过实验发现，在信息加工任务中，通过到长时记忆的词汇信息的特定速度可以解释言语智力的差异，激发了研究者们根据认知加工来估计个体差异性的兴趣。许多人在早期因素取向的基础上从认知心理学的角度着手，开始了参数取向的研究。他们与早期因素取向的主要区别在于，强调的是动态的而不是个体结构的差异，试图用信息加工而不是单独用结构项目来解释个体的认知能力。参数取向的研究者们关注的是认知加工过程中特定参数对智力差异的解释力，假设个体智力表现上的差异源于一些基本的认知能力参数差异。从目前的研究来看，加工速度和工作记忆容量这两个参数得到了认知心理学家们最为广泛的认同。

自从Spearman指出G因素的存在以来，人们对大脑的信息加工机制已有了更多的了解。由于对G因素的解释一直不太成功，研究者希望从脑与认知科学的最新研究成果中寻求支持，其关注的重点首先集中在了信息加工速度或神经效率，特别是大脑如何进行信息加工上。这方面的研究已取得不少重要进展。有许多研究者发现了个体在信息加工速度上的差异，较快的反应（因此推测信息加工较快）与较高的IQ相联系，并推断这可能就是不同个体智力差异的原因。已经有多项研究发现不同年龄的个体信息加工速度不同。如，年长儿童比年幼儿童加工速度更快，老年人比年轻人思考得更慢。也有许多实验发现了不同智力的人群之间信息加工速度的差异，如天才儿童和普通儿童、LD和非LD儿童、阅读良好和阅读不良者在提取速度、思考速度上的差异。此外，研究者发现，无论儿童还是成人，其大多数领域的信息加工速度都与智力存在着有效的相关。斯滕伯格区分了信息加工方面的四种速度：纯速度、选择速度、词汇通达速度以及推理加工速度，发现了加工速度特别是后

两种速度和智力之间可靠的相关。也有实验表明,即使不需要快速反应时,信息加工速度与非限时测验的表现依然有显著的相关。一些文献认为,反应速度和智力至少有中度的相关。这些结果似乎暗示了加工速度和智力之间的关系。由此,研究者推论:提取信息快并且有效的人,在认知任务上会表现出优势,特别是在言语任务上。Hunt 及其同事通过一系列实验任务对这个立场给予了支持,在一个研究中,他们让成人检验关于范畴的陈述,例如:“狗是动物”,结果发现,确认这样的陈述需要的时间与语言能力显著相关。

艾森克认为,生理维的加工速度、解决问题的坚持性、检查核对错误的各个水平组合,构成了智力个体差异的根本。Jensen 认为,有意识的大脑在同一时间内只能处理有限数量的信息,处理的信息在短时间内会很快消退,为了弥补信息加工资源有限和消退的特点,就会求助于复述策略以及长时记忆系统加以保存,然而存储的过程也需要消耗时间,因此心理速度高意味着单位时间内能够处理较多信息,超过认知负载可能性就减少,认知功能就会较高,速度愈快,在特定认知负载情况下其优势就愈明显。由此,也就可以认为,智力个体差异根源性的因素是生理层面的信息加工速度,速度几乎就等同于一般因素 G。

总之,虽然不是信息加工的全部方面,但至少是在大多数方面,都存在加工速度的个别差异,并与智力存在着中等程度的相关,这使得加工速度的参数说更为可信。研究者认为,阅读是科技社会人们最重要的认知技能之一,在提取语词意义速度方面或者把字母整合成有意义的语词方面,一些很小的差异就会导致阅读速度和阅读理解上根本性的差异。因此,尽管加工速度的个体差异很小,常常在毫秒级水平,但很重要,它们在一个微观水平上表明了个体智力的重要差异。

上个世纪末以来,信息加工速度与智力的关系问题再度成为被关注的热点,研究者们以检测时 (inspection time, IT) 作为信息加工速度的指标,并将它与 ERP 研究结合起来,试图从神经生理过程的角度探讨加工速度对智力差异的贡献。目前,已经有许多研究者承认,检测时 (IT) 是最为可靠的信息加工速度指标,是智力差异的基本元素,其基础是神经传导速度;同时,将信息加工速度研究与认知神经科学手段结合起来,已成为当代智力研究的一种新的研究范式。

对于加工速度是智力的重要成分,似乎没有多大疑问,但是有研究者认为,对于基本的认知加工来说,可能还有同样或者更加重要的机制,其作用甚至超过了加工速度,这就是工作记忆。例如,Case 等人通过研究推断出,有较高工作记忆容量的儿童和一般儿童的智力差异主要取决于工作记忆容量的差异。Miller 和 Vernon 的实验给了工作记忆以更大的支持,他们用 WISC 和一系列反应时及工作记忆任务测试了 4 至 6 岁的儿童,结果发现,随着年龄增加,儿童反应更快,工作记忆广度也随着年龄增加,一般反应时和工作记忆存在着显著的相关。不仅如此,Miller 等人还发现,反应时与 IQ 在每个年龄上都有显著相关,但工作记忆和 IQ 的相关更大,甚至在考虑了反应时效应和年龄之后,相关依然显著。这使得 Miller 等人得出结论:对于儿童来说,“记忆解释了智力的显著变异”,而相比之下,反应时则作用较小,至少,在考虑了工作记忆的个别差异时是如此。而 Conway 等人的研究,不仅支持了工作记忆,还否定了加工速度的作用,他在对 120 位大学生的研究中,通过潜变量分析发现,工作记忆和流体智力存在着密切的关系,而加工速度和流体智力之间的相关度就较低。不过,Fry 和 Hale 的研究似乎可以得出不同的结论,他们选取了 200 多名 7 到 19 岁的被试,对每个被试进行了加工速度、工作记忆和流体智力的测验,之后通过分析发现,工作记忆到流体智力存在着比较显著的路径系数,而加工速度到流体智力则不存在显著的路径系数,但是进一步地分析表明,有 97% 的工作记忆的年龄差异是由加工速度的年龄差异引起的,而仅有 3% 的工作记忆年龄差异

是独立的。因此，他们认为工作记忆的发展取决于加工速度的发展。刘昌等对中国儿童的加工速度、工作记忆和液态智力的发展也进行了研究，得出了与 Conway 等人类似的结论，不支持 Fry 和 Hale 的观点。显然，关于工作记忆和加工速度的争论并没有结束，在这个问题上，Kyllonen 的四资源模型试图根据工作记忆容量、信息加工速度、程序性和陈述性知识宽度来解释认知任务操作上的差异，为分析智力的个体差异提供了一种潜在统一的观点，值得关注。

认知机能决定理论认为，G 因素是某一种认知能力，智力个体差异的内在机制就是由注意资源的总量、速度、工作记忆等认知能力的差异导致的；智力个体差异只有认知能力水平的“量”的差异，而没有认知结构的“质”的差异。例如，Earl、Hunt、Lansman 和 Yee 等采用多任务范式研究注意与智力水平的相关，认为智力活动过程需要消耗资源、要对不同来源信息给予监控注意并不断协调；一般智力就是注意，智力个体差异反映的是注意的差异。后亦有研究认为：注意分配与智力联系更为紧密，注意是一般智力的有效预测因子。Vernon、Jensen 和 Jensen 等用检测时任务作为速度指标来描述一般心理能力，认为速度是一般智力，速度是智力水平的有效预测因子。然而，也有研究者认为加工速度不是一般智力的全部，速度与智力的关系会受到其它变量的调节，如任务难度、认知策略、注意调节等。但 Knorr 和 Neubauer 运用复杂的概念检测反应时和一般反应时任务得出的结论都支持了速度说，而没有支持复杂调节说。在对工作记忆影响智力功能个体差异的研究中，工作记忆与一般智力相关稳定，以致有工作记忆就等于 G 因素的说法。Engle 以及 Klaus Oberauer 等认为，工作记忆和一般流体智力反映了干扰和分心情景下保持表征活跃、控制性注意等的的能力。Embretson 认为工作记忆容量与控制加工过程与一般智力之间有稳定相关，控制加工过程比工作记忆容量对一般智力影响更为强大。Schweizer、Mackintosh 和 Schweizer 等认为，维持性注意、工作记忆、知觉速度都是流体智力活动中必不可少因子，维持性注意使得长时信息加工得以连续。Conway 等认为工作记忆与短时记忆是相互独立的结构，速度作为工作记忆的调节因子出现，速度与短时记忆和工作记忆都是智力成就的有效预测变量。

智力的心理速度理论 (Mental speed theories of intelligence) 所关注的是心理计量学智力与认知或生理心理学变量的相关，考察的重点是心理速度，相应的测量指标主要有检测时 (IT)、反应时、ERP 或神经传导速度，等等。在运用 IT 任务进行的 ERP 研究中发现，ERP 波幅和潜伏期的不同往往能够反映出智力上的差异。信息加工速度和神经传导速度均与一般智力因素 (G 因素) 存在较高的相关。上述研究发现促进了智力与心理速度关系的跨学科研究，成为智力研究领域中的热点。在智力测验中，许多测验任务都与被试的加工速度有关，包括在非常简单的作业任务上的加工速度。智力测验所要测量的多项能力都能找到与之对应的速度变量，如表 13.1.2 所示。

表 13.1.2 不同的智力形态及其相对应的速度变量

能力 (Ability)	速度变量 (Speed Variable)
液态智力 (fluid intelligence)	推理速度 (speed of reasoning)
晶态智力 (crystallized intelligence)	阅读速度 (reading speed)
空间能力 (spatial ability)	知觉速度 (perceptual speed)
构思的流畅性 (ideational fluency)	提取速度 (retrieval speed)
认知速度 (cognitive speed)	完成测验的速度 (test-taking speed)
加工速度 (processing speed)	反应时 (reaction time)

另外，像记忆和学习这样的能力虽然没有可明确对应的速度变量，但我们知道工作记忆是和复

述的速度相关的,学习效应则主要体现在知觉加工和概念加工的流畅性上,也都与信息加工的速度有关。因此,心理速度实际上是智力测量的基本指标。

常用的测量信息加工速度的方法是选择反应时(choice reaction time, RT)和检测时(inspection time, IT),其中又以后者最为常见,特别是在考察智力的生物学指标时,研究者通常将智力测验、IT和ERP等多项指标整合在一起使用。

在智力G因素与信息加工速度的生物学研究方面,以美国心理学家Jensen为代表的生物学方向的智力研究把G因素与一些生物指标联系起来,强调速度因素是智力测量的重要依据。Reed和Jensen把G因素归为神经加工速度的指标(an index measure of neural processing speed)。Deary和Crawford运用因素分析法从已发表的各类测验中提取出G因素的载值(G loading values),计算它与一些生物指标的相关。他们发现,G因素与头型、脑葡萄糖代谢率、平均诱发电位的幅值和波形的复杂度,以及视觉诱发电位(Visual evoked potentials)、大脑神经传导速度(Brain nerve conduction velocities)和视觉通路的神经传导速度(Visual pathway nerve conduction velocity,简称NCV)等许多生物指标(biological indices)存在高的正相关。Duncan等人利用PET技术进一步研究G因素与大脑皮层激活区的联系。他们采用了三种认知任务:空间任务、语词任务和感觉运动任务,每一种任务又分为两种水平:高G因素任务和低G因素任务。研究发现,不管是在哪一种认知任务下,高G因素任务伴随的脑激活区和低G因素任务伴随的脑激活区之间的差别都出现在侧额叶(lateral frontal cortex);在高G因素的认知任务中,被试侧额叶的激活水平更高。该实验进一步支持了G因素的生物学存在。

一些研究认为,液态智力(fluid intelligence)与生物学基本特征相关,它随年龄的增长而衰减。与此不同,晶态智力(crystallized intelligence)更多地受到环境和教育因素的影响,并随年龄的发展而变化,与言语能力显著相关。液态智力主要包括推理、记忆和知觉速度,一般用数字测验来测量;而晶态智力包括知识和流畅性,一般用言语测验来测量。根据智力投资理论,液态智力对晶态智力保持动态的影响。液态智力是婴儿期行为的主要决定因素,它会影响文化知识和特定的文化技能的习得。到成人期,这种主导性的影响逐渐转移到晶态智力上。晶态智力随年龄的增长而增长或保持不变,然而基本的生物行为还是主要受液态智力的影响,如早期知觉行为等。

### 13.1.3.3 关于人类个体智力差异的认知策略研究与探讨

策略是旨在实现特定目标的心理操作。主体通过运用策略,可以达到解决问题的目的。自从布鲁纳开创了对认知策略的研究以来,人们逐渐认识到策略对个体认知操作进程的作用。许多研究者都提出策略是智力的核心成分,由此开始了从策略取向角度对智力的个体差异的探讨,目前,主要有两种理论观点。

(1) **策略差异取向**。这种取向假设对于特定问题有多种解决策略,不同的个体在策略的运用上有所不同,从而导致了智力的差异。目前已经积累了有关个体策略差异的多方面证据。策略的发展性研究表明,年龄较大的孩子与年龄较小的孩子相比,运用的策略更为丰富,运用的领域也更为广泛;而老人和年轻人相比所运用的策略也有所不同。对天才儿童和普通儿童、智力迟钝儿童和正常儿童、学习困难儿童和学习正常儿童、学优生和学差生、专家和新手、阅读不良者和同样IQ的阅读良好者等不同人群的研究也表明,策略运用上的差异是其主要的认知差异,这种差异表现在运用的数量、类型、运用效果以及迁移等多个方面。如Gaultney和她的同事在一系列自由回忆实验中,比较了中学阶段的高智儿童和普通儿童记忆策略的运用,发现高智儿童始终比他们的普通同伴更多地

运用了策略。而智力迟缓和正常儿童的比较研究也发现,运用记忆和问题解决策略的倾向是两组儿童主要的认知差异,正常儿童比智力迟钝儿童更可能运用有助于其表现的策略。例如,在记忆研究中,测验前给正常儿童和智力迟钝儿童复述信息的机会,但智力迟钝儿童在准备回忆时比同龄正常儿童更少地进行复述;而且,当要求儿童复述时,可以发现他们在复述上存在质的区别,智力迟钝儿童和智力迟钝成人在复述时,更可能每次只复述一个或两个单独项目,而不是一起复述几个不同项目(累积复述),这样一种简单重复的策略与观察到的年幼正常儿童的策略很相似。总之,无论是在复述的数量还是在复述的类型上,智力迟钝儿童都显示出了复述策略的缺陷,而这直接影响了他们回忆的水平。此外,这种策略缺陷并不仅限于记忆方面,在其他任务和策略上也是如此。

研究者也用这种策略运用上的差异假设来解释学习困难儿童和学习正常儿童在认知任务表现上的差异。比如,Bauer给学习困难的儿童(LD)和正常儿童(non-LD)一系列自由回忆任务,结果发现两类儿童在记忆材料的最后几个项目上没有组间差异,但在回忆最初的项目上则有较大的组间差异。通常人们认为,对那些最后项目的回忆,由于儿童清空了短时记忆储存的内容,不需要任何特殊的策略;而对记忆材料中最初的项目的回忆则反映了策略的应用。Bauer由此认为,LD儿童与正常儿童的差异在于策略的运用而不是非策略的技能。这种解释得到了后来一些实验结果的支持,在实验中观察,LD儿童在运用组群分类这种记忆组织策略时确实水平较低。研究者通过进一步研究提出,不同个体策略的差异不仅表现在策略运用的数量、类型上,还表现在运用的效果上。Gaultney对三、四、五年级LD儿童和正常儿童运用记忆策略的数量进行了统计上的等组研究,发现LD儿童与正常儿童相比,存在着明显的策略利用缺陷;即使有时LD儿童也能像正常儿童一样策略化,但在同等的策略运用情况下,LD儿童却不能像正常儿童一样在记忆表现上从策略中同样受益,Gaultney认为这是利用缺陷的标志。

对策略训练的研究还发现,不同个体在策略学习以及迁移方面,也存在着明显的差异。例如Campion和他的同事训练轻度智力迟钝成人和在智商上匹配的正常儿童解决推理问题,为了解决这种问题,必须学会相应的策略。年幼的正常儿童和智力迟钝成人都学习了如何解决这些问题。然而,当呈现新问题时,发现了组间差异,无论新问题与他们训练的问题相似还是有些新颖性,正常儿童都比智力迟钝成人能够更有效地把新学习的策略迁移到新问题中。由此可见,在策略上的差异是个体智力技能上重要的差异。

(2) **多重策略取向。**随着对策略研究的进展,研究者发现,几乎每个人都运用了多重策略,并且有许多策略是不同人群都会共同使用的。Siegler更通过一系列的研究证明,在几乎所有的认知领域,个体的策略都具有这种特点。由此,研究者认为,以往那种简单的策略差异观点是不正确的,个体在策略上的差异是他们策略运用分布上的不同(例如,在什么时间运用哪一种策略),是这种不同的分布导致了个体的认知差异。例如,年龄较大和年龄较小的孩子都能运用提取(retrieval)策略和数数(counting)策略来计算加法问题,但不同的是,年龄较大的孩子更多地运用了提取策略,特别是在困难的问题上。在一项研究中,Geary和Brown评估了天才、普通、数学困难的三年级和四年级儿童在解决一系列加法问题中对策略的选择和运用,发现所有的儿童都会使用不止一种策略来解决问题,但是天才儿童对多种策略的联合运用最为成熟、最为有效,其次是正常儿童,而数学困难儿童最差。

多重策略取向还指出,虽然有多种策略,但是针对特定问题,个体总会为自己选择最佳的策略;个体间策略选择的不同,可能是因为对特定问题的学习不同,以及那些随着最佳策略而不同的加工

参数的差异而导致的。例如, 年龄较小的孩子对困难的加法问题选择了数数策略, 是因为他们还没有习得足够好的答案, 而且, 如果他们尝试提取策略的话, 可能会出错。相反, 年龄较大的孩子知道答案并会运用提取策略, 因为它需要的心理努力相对较少。

#### 13.1.3.4 关于人类个体智力差异的策略适应性研究

策略适应性取向是研究者对策略取向的进一步发展。策略取向强调个体在策略使用上的差异。随着研究的深入, 一些研究者发现, 不同个体之间在策略运用和策略训练方面存在差异的一个主要原因是元认知; 由于元认知的差异, 不同个体在理解策略的有效性并适应性地运用策略上明显不同; 因此, 这些研究者提出了策略适应性取向。其基本假设是: 个体在其策略选择的适应性上有所不同, 也就是说, 尽管两个个体可能拥有同样的一些策略, 但是, 他们针对特定情境选择最佳策略的能力有所不同, 因而导致了不同的解决问题能力。目前, 这一取向中主要有以下两种倾向。

(1) **元认知知识观点。** Case、Flavell、Sternberg 等人提出, 一些个体缺乏足够的元认知知识, 他们不知道策略为什么有效、也不了解策略在何时有效, 同样, 他们也不知道该怎样运用策略, 因此, 他们的策略运用水平很低, 而这使他们在算术、推理、阅读、记忆等多种认知活动中“事倍功半”甚至失败。有研究表明, 学困生和学优生元认知知识上有很大的不同, 因而导致他们在解决问题策略上存在着质和量的差异, 最后解决问题的效果和能力也大不相同。尽管这个观点似乎很合理, 但是, 从目前来看, 其实实验证据不足。有研究发现, 元认知知识和策略选择的适应性之间只有微弱的关系。而且有反对者提出, 有些个体运用的策略虽然不够复杂, 但那正是他们适应性选择的表现, 这些不够复杂的策略对他们来说可能是最佳的, 因为更复杂的策略虽然高级但需要付出更多的努力, 而且复杂的策略如果缺乏练习还很容易出错。

(2) **元认知调节观点。** 一些研究者主张, 个体不是缺乏特定策略的元认知知识, 而是根据任务转换策略的调节能力不足。在执行任务过程中转换策略的能力是十分关键的, 因为, 许多任务都是动态的, 所以要随着任务的变化改变策略; 另外, 即使任务结构保持稳定, 随着对任务的逐渐熟悉, 个体也必须转换策略的运用。

影响个体转换策略的因素有许多, Reder 和 Schunn 通过研究指出, 其中一个重要的因素就是策略的成功率, 他们采用动态任务研究发现, 人们能够根据策略成功率调节其策略, 而且无论在执行任务的早期还是后期, 个体运用成功率调节策略的能力都存在着系统差异, 而这些个体差异可以很好地预测执行任务的成绩; 他们还发现, 适应性强的被试并不是运用了更为复杂的策略, 而是更善于随着情境选择更适应的策略。并且, 个体在策略是否适应、适应程度以及适应速度方面的差异, 似乎并不能归结为不同的任务知识或者一般的智力差异, 而是独立存在的。另外, 关于问-答和简单问题解决任务中个体差异的分析也表明, 这种策略适应性上的个体差异是一个很普遍的现象, 影响着各种认知任务的完成, 因而应是个体认知差异的重要原因之一。

#### 13.1.3.5 认知发展理论对智力个体差异的认识

认知发展理论对智力个体差异有两对分歧: 纵向的成分关系强弱差异与结构差异分歧, 以皮亚杰认知结构理论和信息加工发展理论为代表; 横向的领域特殊性与领域普遍性差异分歧, 以朴素理论和皮亚杰理论为代表。皮亚杰认为, 智力本质是适应, 智力发展受到遗传、成熟、事物经验、逻辑经验限制, 平衡化机制是智力发展的决定性机制。横向差异是指对同一发展阶段内概念掌握的时间差异, 反映了受心理场和知觉场的线索数目、注意程度、激活的图式是否融入更高层次结构等影响图式活动水平的差异; 而纵向差异是指不同年龄段解决问题时的图式相互协调和逻辑经验导致

的认知结构类型差异。纵向的认知结构差异具有跨领域的一致性。现实化认识风格、形式化认知风格是影响横向和纵向智力个体差异的重要因素。Sternberg把认知发展看做是元成份、操作成份、知识获得成份、保持成份和迁移成份的发展和各个成分之间整合速度的发展,认为智力个体差异的认知机制是各成分分配时间和各成分整合速度的差异。Case认为认知发展是储存和操作两个空间变化过程,储存空间主要是存储知识和策略,操作空间主要是进行信息处理,智力个体差异的认知机制是操作空间和存储空间的知识策略的差异。Sigeler认为认知发展是个体信息加工能力提高的过程,神经系统变化、联想竞争、编码水平、类比策略以及策略选择都会影响信息加工能力的发展。纵向和横向的智力个体差异的认知机制是功能性的,包括各个功能成分的发展和各个功能成分之间联系强弱的变化,而不是结构构成的差异。

Wellman等认为,对心理、物理和生物现象的理解和信念构成了个体的知识系统,这些核心知识有对不同领域事物作本体区分、进行因果解释且能做一致性因果解释推理系统特征。他们倾向于认为认知发展是连续的、领域的,朴素理论差异是智力个体差异的重要来源。

### 13.1.3.6 进化理论和生态文化理论对智力个体差异的认知

文化活动理论认为,活动结构、个性、意识结构、生态文化环境是智力个体差异的重要来源。其中,维果茨基认为,心理机能包括人类进化中获得的低级心理机能和社会文化历史发展中获得的高级心理机能两部分,由作为集体和社会活动中的心理间机能向作为个体活动、内部思维方式的内部心理机能转化是人的心理发展机制;智慧活动经历了活动定向、物质化活动、有声言语、无声外部言语、内部言语等阶段,智力活动还受到态度、性格等的影响;最近发展区是个体真正的智力发展水平,智力个体差异是最近发展区差异,包括个别差异和情境差异,个别差异是不同个体之间最近发展区不同;情景差异是同一个体在不同情景可能有不同的最近发展区。语言、社会活动结构、智慧活动阶段的质量、个性等是影响智力个体差异的重要因素。Berry认为,人类个体在接受生态影响、遗传影响和文化传播之后才产生可观察的行为和可推测的行为心理特征,其中生态力量是文化和行为的原动力;一定的文化模式塑造和养育着人的心理和行为。包括自然栖息地的生态学环境、包括个体学习和成长的背景的经验性环境、影响特定行为的操作性环境、人为设置的情景性实验环境等,是影响智力个体差异的四个层次环境类型。智力心理活动中的理解和适应性训练,受生态文化环境的多重影响,智力个体差异具有文化相对性。这些观点得到了文化人类学研究的支持。

### 13.1.3.7 基于神经活动效能的智力个体差异解说

智力个体差异有着大脑工作效率差异的基础。神经加工特性论者认为,智力个体差异源于神经加工特性水平差异。比如,Garlick认为,神经网络可塑性就是智力的一般因素,是可塑性程度差异导致了神经网络的连接程度和模块化速度差异。Schafer的神经适应理论认为,高效率的大脑处理熟悉刺激时会使用较少的神经元群,处理较陌生的新异刺激时会使用较多的神经元群;高智个体有较快的习惯化过程和去习惯化的过程,低智个体需要较长时间建立习惯化过程和去习惯化;智力个体差异主要是神经暂时联系过程时相和空间差异。Robison认为,信息加工过程必须处于适度唤醒状态,皮层下唤醒水平是制约信息加工质量的重要因素;反应性和坚持性是唤醒能力的两个指标,高反应性和低坚持性的皮质唤醒活动有助于液态智力成分的活动,低反应性和高坚持性有助于概念思维活动,而中等反应性和中等坚持性有助于记忆加工。一般来说,反应性影响对具体事物的初级和次级表征,而坚持性会影响知觉和概念的形成,智力个体差异有唤醒能力差异的基础。吴天敏提出,智力是脑神经活动的针对性、广扩性、深入性和灵活性在任何一项神经活动和由它引起并与它

相互作用的意识性的心理活动中的协调反映；这四种特性又是可塑的，其可塑性是以神经机能的有限可塑性为基础的；这四种特性是神经活动中神经联系的方向性、扩散和集中、心理和生理相互作用的灵活性的反应。他认为，智力个体差异是四种特性的综合作用，灵活性是智力个体差异的更为重要的指标。

脑区加工特性论者认为，智力个体差异反映了特定脑区加工效率的差异。很多研究者认为，个体越聪明，那么他与任务相关的脑区的工作越高效；而智力程度较低者的任务相关脑区，不仅低效，而且很多与任务无关的脑区也处于工作状态。聪明者具有较为高效的皮质工作定位，结果具有较少的脑活动；智力较低者的神经组织的组合较少且不高效率。这种观点更倾向于认为大脑的机能区域分化和整合水平决定了智力个体差异，而不是单纯的神经信息加工速度。

### 13.1.3.8 人格分化差异理论对智力个体差异的认知

人格分化差异理论认为，智力功能水平和智力结构内部状态受认知风格、自我、人格特征的影响，这种影响可能是通过人格中介特定情景来影响智力的，也可能是人格在各种情景下普遍限制了智力的功能。

智力个体差异的研究者对智力个体差异与认知风格的关系的认识并不一致。比奈非常重视错误答案和正确答案背后的主观/客观、意识/无意识、科学/文学的思维风格，认为正是这些体现了“质”的智力个体差异。Witkin 等认为认知风格是知觉和智力活动过程中表现出来的具有自我一致的机能模式特征，认知风格与智力存在稳定的相关关系，认知风格影响着流体智力操作成就，场独立型个体倾向于取得较好的瑞文推理成绩。但也有研究认为认知风格与智力活动成就不存在相关。有的研究者则认为认知风格与智力之间的关系受问题类型的制约：人际社会问题有利于场依存型个体的智力表现，而非人际社会问题则有利于场独立型个体的智力表现。

对智力个体差异与人格类型的相关研究发现：神经质与智力、特殊能力呈负相关。这意味着个体神经质水平越高，则在测验情景中将具有更多的焦虑成分，焦虑则导致个体远离解决问题活动，导向与任务无关的活动。神经质的影响是整体性的。智力成就与外向型呈正相关，外向型个体的皮质静息状态水平较低，能够迅速启动；而内向型个体的静息水平较高，导致个体从静息到动作电位的转化水平较高；外向型程度高的个体在测验长度较短、限时测验中会有较高的皮质激活优势，而高精神质的个体则一般具有较高的创造性。

智力与人格特质的相关研究发现：开放性能够部分预测言语因素。开放性较高的个体会倾向于卷入言语和文化特点的智力活动，进而就发展了言语推理能力；开放性和智力成就呈中等相关，开放性和流体智力是相互独立的结构，但是二者联合起来决定了知识。外向型与智力成就相关方向受测验情景和测验类型调节，宜人性、责任心与智力呈负相关或者无相关，智力水平越低，则个体会在适应中会投入更多的努力。

Demetriou 提出了一个智力与人格并行层次模型：一层为领域特殊性的认知模块与人格因素，前者引导着信息加工系统活动的领域，后者决定着个体在社会-环境水平的互动关系模式；二层为认知过程和气质，前者决定信息加工的复杂性和信息类型，后者决定了信息如何接受如何反映；三层为目标设定、加工资源分配、难度评价、高层认知、自我防御、工作的自我概念等，是关于加工过程的自我监控和关于自我价值感的自我监控；四层为认知效能与自我价值感，是关于认知效能和自我价值的认识；五层为自我表征的认知模块与人格因子的自我表征，是关于各个领域特殊性的认知和人格因子的自我认识；六层为共有的思维风格。他发现在相同层次之间存在着相关关系，但是在

非同层之间的相关关系较弱，高层制约低层。

#### 13.1.4 智力个体差异研究展望

多年来，众多的研究者在智力差异研究领域做出了卓有成效的工作，众多的智力个体差异研究表明，智力是一个多层次、多维度、多形态、多要素的介于环境刺激和反应之间的中介性调节机能。从要素上看，它包括认知能力和非认知能力成份；从形态上看，它包括着非知识形态 and 知识形态的智力成份；从层次上来看，它包括着非意识层面和意识层面的智力成份。造成智力个体差异的因素是多样的，既有智力结构内部的原因，也有智力结构外部的原因；从智力结构来看，既有要素水平的，也有要素关系水平的，还有要素结构性质的；从智力个体差异的来源上看，主要有社会因素、经历因素、教育因素和遗传因素等。

因素取向研究从传统的心理测量学理论出发，对智力个体差异进行了开创性的研究，但它探讨的是个体之间在静态结构上的差异，忽视了个体智力的活动性。因此，不少研究者批评基于因素取向的 IQ 测验虽然能够评估个体智力的“差异”，但却不能成功洞察智力差异的本质。

出现智力测验预测效度较低的现象，是由于智力的本质尚未被揭示出来的缘故。显然，智力是一个较难穷尽的多元组合结构。有人认为，在智力的各种外围成分上作过多的争论，对于智力研究的发展是没有意义的，它只会使这一研究陷入思辩的泥潭。一般智力才是智力的真正实质。也许重新回到斯皮尔曼的 G 因素，重视它的科学和社会启示作用，才更有益于智力及其测评的研究和进展。

基于因素取向的智力测验偏重于对一般智力的测定，那么，一般智力又是什么呢？正如 Spearman 发现的那样，若对各种认知任务的成绩或成就进行测量，那么测验间的相关通常是正的。在一种测验中成绩较好，至少在一定程度上，在另外一种测验中也倾向于做得较好。为了解释这一结果，Spearman 才提出了一般智力或称 G 因素的假设，认为它有助于人顺利完成不同的任务。另一个假设由 Godfrey Thomson 提出。他认为，任何一个任务都包含一系列的成分因素或信息加工功能，测验间普遍存在的正相关不是由于某一种共同的原因，而仅仅是因为任何两个任务可能至少包含一些共同的成分。所谓的“G 因素”其实就是整个认知功能的平均效能。由于缺乏因果资料，仅依靠相关资料，由此而导致的因素分析的不确定性使得我们还很难证实哪一种假设。当然，也不是所有人都认同上述两种对一般智力的看法。随着现代智力理论和现代智力测量理论的发展，人们发现一般智力的范围似乎要超过以上这些假设。智力的三元理论就认为，一般智力应包括分析性智力(传统的一般智力概念)、创造性智力和实践性智力，而且它们是相互分离的。该理论认为，这三种智力在不同文化背景、不同年龄和不同社会经济地位的人身上是不一样的。传统的因素分析研究之所以能发现一个 G 因素，是因为它们采用的测验的范围过窄。

后来的参数取向研究是因素取向研究的延伸，它与因素取向的主要区别在于它从信息加工观点出发，主张从信息加工观点描述发展机能的机制同样可以用来描述个体智力的差异，因此，信息如何编码、加工速度、信息如何分类等方面的差异等都可能成为智力差异的来源。在参数取向的研究者们看来，即使个体策略的选择也是一个过程，可以被认知结构上的参数差异诸如推理能力、工作记忆容量和加工速度等所影响。该取向试图寻求超越特定领域任务的个体预测特征，解释跨越多种任务和领域的有关的成绩差异。

策略取向研究认为，信息加工速度和工作记忆显然是智力的重要成分，然而，它们只是基本水平的加工，还有更为宏观的加工。个体在高级水平认知加工上的差异影响着个体的智力差异，其中之一即是策略。该取向着眼于分析策略，认为策略是完成任务的根本要件，它使人能够计划认知操

作的进程、预期他们与他人行为的结果，是智力的核心成分。因此，策略取向的研究者们尝试从策略差异的角度对个体特定任务表现提供详细的解释。以往，在参数取向与策略取向之间，两派研究者往往强调其中一个而否定另一个，但实际上二者并不一定是非此即彼的对立关系。已有研究表明，不同群体会选择不同的策略来弥补参数的差异，例如，老人会更多地依赖于看似有理的推理策略，因为他们提取精确记忆要比年轻人付出更多的努力。

策略适应性取向研究从策略适应性差异角度对智力和任务成绩之间的关系进行解释，力求在结合参数差异取向和策略差异取向优点的基础上，洞察策略选择机制和个体差异性质。虽然这一点尚未得到一致的认同，比如，Siegler 等人就指出，对以前研究表明存在加工速度和工作记忆容量差异的老人和年轻人的比较中，没有发现策略适应性上的组间差异。但 Reder 和 Schunn 认为这可能是由于 Sieglar 等人把适应性看成了一个静态的领域。可以说，至少适应-差异取向在参数取向和策略差异取向之间提供了一种联系，即选择不同的策略是因为不同的策略适应性，而这与参数差异有关。不过，策略适应性取向也面临着质疑。多重策略取向就指出，每一个人在他们的策略选择上都是适应的，都能为他们选择最适合的策略，各种人群策略选择的适应性并没有显著差异。例如，年轻人和老人、不同背景的儿童，策略选择同样适应。另外，对策略适应性和个体能力关系的研究发现，工作记忆能力能够预测个体对情境觉察的速度以及个体适应速度；而推理能力能够预测个体在特定情境下对最佳策略的理解，这不免使人感觉又重新回到了参数取向。

上述研究无疑是值得肯定的，但是，这些已有的研究仍存在有诸多的不足。比如：整群研究有余，而对不同智力水平群体的差异机制研究不足；不同的研究取向都存在着各自的欠缺；很多研究的被试多取自学校的不同智力水平的学生群体，不加区分的整群相关研究的结论只能解释中常整群的智力水平差异中较为重要因素，而群体间智力功能的差异可能是由于能力要素水平和能力结构形态不同导致的。因此，还需要对不同水平群体的智力结构要素水平、各要素水平形态结构、智力活动的基本特点进行研究，以揭示不同智力水平群体的智力活动特有的机制。更需要构建比较全面的智力差异理论，从认知加工和神经机制研究相结合的角度来恰当地解释智力差异，同时加强对非认知因素的关注。

在智力与人格的相关研究中，不少研究忽略了不同智力成分与其它系统的非线性事实。神经水平的表征、映像表征、表象表征、言语表征等，应是智力活动必不可少的表征层次。从加工对象来看，要涉及到对外部信息的加工和对加工过程的自我意识的加工；从意识层次来看，既有有意识的控制加工也有无意识的非控制加工的卷入；从加工背景来看，人格以其潜在的、稳定的心理活动模式制约着智力活动过程中对外部信息的开放状态、情感反应模式、加工方式、自我资源投入的多少。智力表征活动总是需要一定的意志和情绪的参入的，它们对智力表征效果的影响存在着人格中非自我部分以神经活动状态影响神经表征层面的智力活动效果、人格中的自我层面则以自我资源的多少来影响智力活动过程中的持久性和唤醒程度、习惯化的认知方式又制约着表征过程的组织等多种途径影响的可能。可以说，智力加工效果受多重交互作用影响，各种影响因素以系统的整体方式影响着智力表征效果，而非线性叠加。

智力水平差异有质和量两个方面。现有研究对不同年龄群体的智力个体差异机制并不区分智力要素功能特征和智力结构特征，因此，应该在研究过程中在区分智力个体差异的性质的基础上整合分析研究结果。另外，静态研究并不足以揭示情景性和历时性的智力结构要素和人格要素系统之间的动态关系。现有人格与智力的相关研究多是特质水平的，而人格特质的形成要经历一系列情景过

程,那么,可以推测,智力个体差异的形成也经历了情景特质和情景下智力活动过程的相互作用。因此,总体的、一般的、不区分情景的人格与智力个体差异不相关并不意味着二者没有相关关系,只不过可能是因为其中还具有其它中介变量的影响,以致二者还没到显著相关的程度。

根据智力个体差异研究的不足以及对智力、智力个体差异的理解,有人提出,未来的研究应在下述几个方面加强:

一是要注重在区分智力形态与子类的基础上研究影响常态群体智力个体差异的因素,在整合差异因素的基础上建立智力结构模型。区分智力形态与子类基础上的因素影响研究一方面是研究认知能力结构内各个操作能力水平个体差异、各个认知操作能力的组织特征个体差异,另一方面是研究个体的动机、兴趣类型、人格特点与认知能力之间的关系,从而揭示个性倾向和人格等对智力个体差异的影响机制。从系统功能与结构关系来看,智力结构的研究不仅应关注智力要素水平的差异,而且还应关注要素关系密切程度、关系性质的差异。智力不单纯是认知能力,智力只不过是整体人格的外在指标。将影响智力差异的其它心理因素纳入智力结构模型是完善个性意义的智力的应有之义;因此,在整合差异机制基础上建立不同智力水平群体的智力结构模型,也是智力个体差异的应有之意。

二是要注重把生理基础和社会基础研究同智力生成过程研究相结合,从来源上揭示智力个体差异的产生机制。智力活动受社会、文化、历史和个体活动结构的影响,智力活动阶段和活动结构有其社会基础。个体间和个体内的智力活动是以脑为基础的,在遗传物质、生化、神经、机能分区、大脑半球基础上研究各种影响因素和智力活动过程之间的相互作用,可以更好的揭示智力个体差异的本质和智力本质。

三是在智力心理活动中需要引入心理动力和人本等概念。传统智力个体差异研究大多将智力与其它心理系统视为正交,仅从认知方面研究造成智力个体差异的智力因素。但心理是整体的、有精神动力的、有精神导向的。有研究表明,宿命论、自信对认知机能老化有较强的预测作用,成人的推理成绩受到与推理内容有关的个体信念的影响,自我损耗后以及与自我损耗有关的态度变化会降低个体的流体性智力活动成就。可以说,国民心态、民风、个人的信仰等社会意识因素以及意志也是造成智力个体差异的内外部因素。

将个性和共性的智力结构相结合,重点和普遍相结合,认知和非认知因素相结合也是智力个体差异的重要研究趋向。智力存在于各种表征形态的适应性活动中,体现了智力的个性;而基本表征过程则体现了智力的共性。共性的智力离不开不同表征内容、表征过程的支持。智力的发展是普遍和重点发展的结合,不同发展阶段有某些重点发展内容。对智力个体差异的研究应该既要重视内容对发展的制约,也要重视功能要素对智力发展的突出作用。智力处于多种主、客观关系之中。非认知因素既可以潜在地影响智力的认知活动,对非认知因素的认识和调节也可影响智力活动。具体来说,可以将知识表征、能力的发展与智力活动的基本过程的心理品质研究相结合,从年龄发展中比较智力活动的心理品质的组织结构特点,从而了解智力个体差异中质的差异;在受控的智力水平下,观察不同心理状态下、人格类型下的智力活动效果等。总之,应该采取多么变量的研究范式,并辩证地把相应的研究结果以系统的思维加以整合。

在未来的研究中,我们期待智力差异研究能从以下方面取得突破:(1)理论的整合。已有的理论还不能对智力的个体差异做出比较全面的解释,它们各自强调了不同的侧面,但却缺乏有机的整合,因而从一个更高的层面建构智力差异理论的框架,合理地解释智力差异现象,进而为智力开发、

落后人群干预提供理论依据将是一个重要的任务。(2) 研究手段的突破。随着大脑信息加工机制研究的进展,在智力差异研究中,将认知加工和神经机制两个层面结合已经成为可能,比如关于信息加工速度与智力的研究已经重新成为了当今的热点,在研究手段上采用了将 ERP 技术、智力测验和 IT 实验等结合起来的方法,并取得了重要的进展。因此,如何从心理学的机制上解释 ERP 实验获得的那些技术参数与智力水平、IT 的关系,确认 IT 在智力差异上的意义,将是摆在研究者面前的一个新的课题。(3) 认知因素与非认知因素的综合。尽管在一些新的智力理论中,已经强调了非认知因素的作用,如情绪智能理论、Ackerman 的 PPIK (Process, Personality, Interest, and Knowledge) 理论,但在对智力差异的解释中,以往对非认知领域的关注仍显不足。有研究者提出,信息加工速度之所以与 IQ 有较高的相关,是因为人格因素的作用,因此,已经有不少研究者开始注意到人格与智力的关系并展开研究。比如,研究者已经发现智力与神经质 (Neuroticism) 负相关、与对经验的开放性 (Openness to Experience) 正相关、与尽责性 (Conscientiousness) 负相关等,Chamorro-Premuzic 和 Furnham 等人还在整合已有研究的基础上提出了人格与智力交互作用的模型。种种迹象表明,将智力差异与人格等非认知因素结合起来,将会是未来智力研究领域的一个新的热点问题。

### 13.2 关于人类个体智能差异测评的研究

如何衡量和评价一个人的智力素质和智能发展水平,对于有针对性的开发和提升一个人的智能水平有着非常重要的意义。有了正确的认知和评价,也就有了发展、调整和提高的方向。

在人类进化的历史进程中,智力的进化无疑是一个重要的方面。由于高智力的人会对环境有良好的适应能力,因此,在自然选择过程中,高智力的因素就会被保留下来,并得到发展。人类发展的历史业已证明,在整体上,现代人类要比远古时期的人类聪明许多。在达尔文思想的激励下,达尔文的表兄弟高尔顿 (Francis Galton) 第一次对智力进行了系统性研究。他认为,智力水平的高低与神经系统的完整性和功能有效性有关,外部世界的信息是通过我们的感觉到达于大脑的。感觉越敏锐,获得的信息越多;信息越多,判断与思维就越有用武之地。高尔顿认为,感觉辨别力“基本上是心智能力中最高能力”。

世界上第一个正式的智力测验,是由法国心理学家比奈和医生西蒙在 1905 年编制的。比奈认为,智力是由多种能力组成的,智力测验必须包含大量不同类型的测验项目。同时,年龄也是影响测验结果的一个关键因素,最聪明的 3 岁的孩子也不能与一个智力一般的 9 岁的孩子相提并论。尽管对智力测验的争论一直没有停止过,但即使是对智力测验持异议的人也承认,智力测评理论和方法的研究与发展,对于培养、改善和提升人们的智力可发挥独特的作用。

在心理学中,智力通常被认为是个体对客观事物的认识能力,是各种认识能力的总和,是获得知识的能力,是认识、理解事物和运用知识与经验解决问题的能力。智力商数 (Intelligence Quotient IQ, 简称智商) 是用来反映一个人智力水平的比值,它体现了个体的观察力、注意力、记忆力、思维力、想象力等方面的综合能力。智商 (IQ) 概念最早由德国心理学家斯腾 (LW Stern) 提出。1905 年,在高尔顿 (F Galton)、卡特尔 (J. M. Cattell)、艾宾浩斯 (Ebbinghaus)、斯皮尔曼和斯腾等人研究的基础上,比奈发表了《诊断异常儿童智力的新方法》一文,成为世界上第一个提出衡量人类智力程度的理论和方法的心理学家,并制订了全世界第一个智力测验表——“比奈-西蒙量表”。美国心理学家推孟 (L. M. Terman) 对比奈的研究成果进行了重要改进,并首次形成了后来广为使用的

“斯坦福-比奈智力测验”和“斯坦福-比奈量表(Stanford-Binet scale)”。在智力测评的发展史上,特别是在现代发展历程中,智商的测算方法主要是两大类。一类是以推孟在比奈-西蒙量表的基础改进而成的方法,称为“比率智商”,即“智商等于智力年龄除以实际年龄”(IQ = MA(智年龄)/CA(实际年龄)×100)。这是一种与常模比较的相对年龄比率分数。另一类是由美国心理学家韦克斯勒(D. Wechsler)在“韦克斯勒智力测验”中首创的“离差智商”,即“以每一年龄段内全体人的智力分布作为正态分布,以个体在这一年龄段分布中距离均数的位置,判定个体的标准分数。”这是一种与常模比较的相对位置比率分数。

对人类个体智能[包括智力、能力和相关心理素质]发展水平和个体差异的测评研究主要包括两个方面:一是人类个体智能的发展水平和个体差异能不能测评,二是如何进行科学而准确的测评。

一种意见认为,智能是对一个人“智”或“能”的度量。是对一个人在面对某事、某物、某种状态或现象时反应的速度、灵敏度以及处理的合理性的度量。有人主张对智能的测评应侧重于“智”的方面,认为智能的测评主要应是对一个人聪明程度的测评,若一个人头脑灵活、理解快、点子多,则认为他聪明、智商高;有的人则主张智能的测评应侧重于“能”的方面,认为智能的测评主要是对一个人能力的测评,若一个人办事能力强、效率高、则认为他能干、能力强。不管是那一种主张,实际上人们都承认,智能是可以测评的。

实际上,以“智”“愚”来品鉴不同个体的观念自古就有之。而力图采用科学的方法来评定人类个体智力高低的历史,迄今也已有一百多年。1905年,比奈(Alfred Binet)与西蒙(Theodore Simon)受法国政府之委托所编制出的比-西量表(Binet-Simon Scale),曾是世界上的第一个智力测验标准,其目的是预测可能会有学习困难的学生。随后,智力测验曾被世界各国广泛地使用,智商(IQ)也逐渐成为了家喻户晓的可度量人们智力高下的重要概念。虽然,这种传统的智力测验方法满足了人们对人类个体智能测评上的便利,但也由于其理论与测量方法的偏失而潜在地制造出不少更严重的问题。例如,传统的智力测验通常仅将人类智力的范围局限在语言与逻辑方面,并假定人类个体的特质能被单一的、标准的、可量化的数据所描述,如此的“IQ”式思维,不仅使得人们近乎被“洗脑”式地被忽略了人类智能的多样性和多元的发展,造成许多具有不同天赋和才能的学生受到了不应有的贬抑与忽视,同时也合理化了传统的“制式”的教育方式,认为可以采用单一的课程和教学方法,以及“标准化”的测验来对待所有学生,而忽视了人类智能多样性的重要性。我们认为,传统智力测验出现的问题并不是说人类个体智力发展水平和个体差异的测评没有必要,而是说,人类个体智能发展水平和个体差异的测评方法必须改进,并且测评的目的和评价观也必须重新认识才行。

### 13.2.1 人类个体智能测评的基本原则

#### 1. 智能的测评要有正确的目的和评价观

对人类个体智能发展水平和个体差异的测评,首先要有明确的目的和正确的评价观。我们进行智能差异测评的目的可能有多种,但就总的来说,我们认为,对人类个体智能的发展水平和个体差异进行测评,其主要目的应是:论证个体智能现有发展水平和个体差异的客观性;预测或发现个人智能发展的多种潜能和可能性;为个人或团体“有的放矢”的提供一些带有指导意义的建设性意见;帮助人们准确把握自己智力和能力的强项和弱项,并根据自己的具体的智力、能力和现有素质状况,确定未来的发展方向,以便轻松愉快地去迎接正在到来的竞争时代;让每个人都有自己的发展空间,使其可以为人类的物质文明建设和精神文明建设,更全面地发挥出自己的聪明才智。若具

体来说,也可认为主要有如下一些目的:(1)为选拔各类人才提供服务;(2)为优化和发展个人的智能结构提供服务;(3)为培养人才,推进合适的个性化教育提供服务。

在所有的智能发展水平和个体差异的测评活动中,对儿童智力发展水平和个体差异的测评也许是人们最感兴趣的。我们认为,对儿童的智力发展状况进行测度和评价,应以帮助儿童发展为主要目的。在对儿童智力的传统测评活动中,过去的某些方法过分强调其甄别和选拔功能,而忽视了智力测评的发展功能。测度者往往是花大量的时间来对儿童进行排序、评定等级和“贴标签”等,即主要是为了选拔适合不同学业教育的儿童服务,而不是为了儿童的未来发展服务,也不是为了创造出更适合儿童的教育方法服务。其最直接的危害,就是使大多数儿童都成为了这些测评方法的受害者,使他们的自尊心和自信心受到了某种挫伤,使其内在的潜能得不到发掘,使其发展受到阻碍,甚至对儿童的一生都产生消极的影响。因此,我们应改变传统儿童智力测评的目的观,将儿童智能测评的目的转变到促进儿童发展这一方向上来。

要帮助儿童首先就要了解儿童,智能发展水平和个体差异测评的目的之一就是要识别出儿童智力素质构成的强项和弱项,进而能在充分肯定其智力素质强项的基础上,针对其弱项采取一些有效的补救措施,扬长补短,以达到促进儿童发展的目的。我们认为,识别出儿童智力素质构成中的强项和弱项并不是智力测评的最终目的,而是希望在此基础上关注并赞扬儿童的智力强项,并以此为切入点,相应地给儿童提供一些适当的引导、恰当的学习经验和更多的表现机会,以促进其发展。

在智能测评中,如何对人类个体的智能状况做出科学而又准确的评价始终是个问题。我们坚持认为:人的智能是多方面的,智能的表现形式也是多种多样的,因此,对聪明或有能力的评判标准也应该是多样的。不同的人,其智力或能力是不同的,而且是可变的和可发展的,这主要取决于两个方面:一是先天的遗传素质;二是后天的学习、实践和教育。一个人先天所获得的智慧特征,只是决定他今后智慧类型的基础;一个人后天所形成的智慧特征,才是决定他今后智慧类型的根本。这也就是说,先天的智慧特征,并不能完全决定后天智慧的类型与发展。因此,认识到每种智慧类型在未来的社会生活中都有可能发挥出其独特的作用,而且认识到每个人都有各自的能力强项,是我们的基本出发点。我们应该说每个人都是聪明的,只是他们聪明的领域和程度有所不同,特别是对正在发育中的儿童更应如此看待。学校教育的宗旨主要是为了启迪和开发儿童的多种智能并帮助他们发现适合其智能特点的职业和业余爱好,因此,对儿童智力的评估应是一个获得其智力发展现状和各种潜在能力等信息的过程。其希望达到的目的:一是帮助儿童,为其提供有益的反馈;二是为该个体的成长提供合适的环境。现有的智力评价体系,在对儿童的传统评价中,过分强调评价的甄别功能,而忽视评价的发展功能,不是为儿童的发展服务,不是为了创造适合儿童的教育环境,因而并不利于其未来的发展。因此,我们应改变传统的评价观和目的观,采用更科学的评价观,使评价始终围绕着促进儿童发展这个目的来进行。

## 2. 智能测评的测评方法和内容要多样化

尽管我们承认人类智能的个体发展水平和差异是可以测评的,但如何准确测评一个人智力和能力的水平还是一个难题。比如,有人主张以一个人在某一活动中的效率方面的差异来测定一个人的智力或能力,认为活动效率就是指活动成绩与活动时间的比率,人们在学习、工作或社会活动中所表现出来的能力的差别就是对一个人智能水平的测度。但是,造成一个人在某一活动中的效率方面的差异的因素通常是多方面的,比如,个人的身体健康状况,个人对此活动的兴趣或动机的强弱,个人的与此活动有关的经验、知识的多少或技能的高低,其内在的智力或能力等。因此,在没有排

除外界因素影响的情况下，所测得的一个人的智力或能力都是不准确的。也就是说，从理论上讲，我们只有在两个人“同样健康、同样努力、掌握有同样多的相关知识和技能”，但对某一活动的效率依然不同时，才可认为两个人在智力或能力方面确实存在差异。但是，我们对其他因素的影响如何才能完全排除呢？这是我们在进行智能测度时始终要考虑的一个问题。

我们认为，由于智能是一个广义的概念，对智能的个体发展水平和差异进行测评的对象不尽相同，智能测度的目的也不尽一致，因此，智能差异测评的方法和内容也不应完全一致，可以多样化。

智能差异的测评方法是指用于测定和评价被测对象智能差异的各种方法。我们提倡智能差异的测评要根据不同的测评对象和不同的测评目的而选用不同的智能差异测评方法，并且要尽量以人类的各种实践活动为基础。智能差异测评的内容是指测评活动所关注的被测对象的某些具体方面或要素。确定测评内容是测评活动的一项核心工作，它对测评活动的开展和评价的准确性有重要影响。儿童智力发展的表现是多方面的，社会所需要的人才也是多样的，因此，无论是儿童还是对成人，智能差异测评的方法和内容都必须是多样化的，应把与测评目标相关的各种能力和智能的测评都毫无偏见地纳入到评价体系中去，使差异的测评可以对被测对象的智能构成能得出正确的评价。

加德纳的多元智能理论认为，人有多种基本智能，每种基本智能又都包含有更次级的多种智能要素，它们在人的智能结构中都占有重要的位置，处于同等重要的地位，在人们的日常生活中发挥着各自独特的作用。事实上，几乎任何文化背景的人，都需要运用多种智能的组合来解决问题。例如，一个小提琴手并非仅靠音乐智能就能成功，他的成功还需要体育运动的高难度技巧、良好的人际协作关系智能以及自我认识智能等。因此，任何一种智能对于解决问题、创造产品都是有价值的。我们在确定评价内容时要充分认识到多种智能在人的智能结构中的同等重要地位，以及在个体未来社会生活中的独特作用，从而才能毫无偏见地把它们都纳入到要评价的内容之中。

我们赞同把能力、情感态度和知识也纳入评价内容之中，并突出强调对解决问题和创新的实际能力的评价。认为解决问题和创新的能力应是人类智能的核心内容。情感态度即个体对待不同领域问题和对象的不同的兴趣、自信心水平、坚持性水平以及对细节的关注程度等。这些在人们的学习和日常生活中常起着重要的作用。传统的智力测评关注的主要是人的知识和认知能力，如识记和再现的清晰度和准确性等，忽视了对个体获得知识过程中的情感态度与价值观的评价，这是不全面的。尽管评价的内容不可能做到面面俱到，对知识的掌握程度进行评价也是必要的，但若忽视了对情感态度的评价，起码是不全面的。在智商测评的基础上，一些人曾提出应加强对情商的测评，这本质上就是对以往错误的一种纠正。

### 3. 智能的评价要多元化和社会化

智能测评的评价标准是指评价时预先确定的对被测对象的某些方面或要素进行评价的度量标准。它们是确定评价对象的某些方面或要素发生变化程度的尺度。评价标准制定的是否合理，是否具有可操作性，是否适合所有被测人群，将直接关系到测评的有效性和科学性。由于每个人的智能都有其独特的表现形式，每一种智能也有多种表现形式，因此，不可能存在适合于所有人的统一的评价标准。因此，提出多元的、弹性的、有针对性的评价标准是一种必然的选择。

传统的智力测验往往把儿童置于一个预先确定的标准或模式[常模]之下，将儿童多样的个性和多彩的表现强行纳入一个由成人预先确定和控制的框架之中，以“一元化”的评价标准对儿童进行评定，忽视了儿童的个体差异，忽视了个体智能构成的丰富性和复杂性，不利于促进儿童富有个性的未来发展。因此，在测评中，我们要抛弃这种“一元化”的评价标准，制定出“多元化”的评价

标准。传统的智能观对儿童或人才的评价,只重视认知能力,内容片面。评价标准也存在单一化的缺陷,只重视统一要求,忽视个别差异。单一的评价内容和单一的评价标准不利于培养和选拔社会所需的多方面的人才。这也就意味着,要进行正确的智能测评,我们首先要变革传统的聪明观,扩大聪明的范围,认识到智能是多方面的,智能的表现形式是多种多样的,因而聪明的标准也应该是多种多样的。甚至可以认为每个人都是聪明的,只是他们聪明的领域和程度各有差异。

除了制定个性化、多样化的评价标准,在对儿童智力的测度中,还要制定出科学的评价标准。由于每个儿童的智力发展至少有两种发展水平:一是现有的发展水平,二是在有指导的情况下,通过自己的能力和别人的帮助所能达到的水平。我们在对儿童进行评价时,要尽量在儿童发展的现实性和可能性之间制定出一个适合的标准,以评价每一位儿童是否在各自原有的基础上有所发展,这既是评价标准多元化的体现,也符合现代教育所倡导的“纵向评价”的要求。它既指出了儿童的个别差异,又激发了儿童的成就感和自我满足感。

我们认为,由于遗传和环境的交互作用,每个儿童都是具有独特个性的个体,他们具有不同的心理倾向、智力特点、兴趣爱好和学习风格等。这些智能方面的差异决定了这些儿童看问题的方法和解决问题方式不可能完全相同。对于同一个问题,有着不同智力强项的儿童的看法或解决问题的方式可能是不一样的,因此,对儿童发展的评价必须以儿童的个性特点为依据,尊重儿童的个体差异,制定出个性化的、多样化的评价标准。对于儿童来说,多元化的评价将有利于培养儿童丰富的个性和创造性,有利于促进儿童富有个性地全面发展。也只有这样,才能在评价中找回大部分儿童的自信,使他们在自信中获得进步,在进步中得到更大的发展。

对成人智能的测度由于目的和要求不同,智能测度的评价标准更不可能一致。依据测度的目的和要求,制定出一套适合的评价标准,当然也是十分重要的。由于对成人能力测度的目的事先都比较明确,因此,如何根据测度的目的制定出一套适合的测度方法和评价标准就成为了人们研究的重点。另外,对人才和智能的评价原本就是一项社会评价,根据社会的认知和需要对评价标准作动态的调整,应是天经地义的事情。

#### 4. 智能测评主要是要设法识别出不同个体智力或能力的强项和弱项

多元智能理论认为,人有多种类型的基本智能,这些不同类型的基本智能都是相对独立的,各自有着不同的发展规律并使用不同的符号系统;各种相对独立的基本智能以不同的方式和程度有机地结合在一起,使得每个人的智能结构各具特点;同一种类型的基本智能在每个个体身上的表现形式也是不一样的,每一种基本智能在人类认识世界和改造世界的过程中都发挥着独特的作用,而且具有同等的重要性。人在解决问题或创新时拥有处理数种不同符号系统的潜能,因此每项基本智能均应遵照各项潜能的核心内涵和发展规律作实际评估。

研究认为,一个人的智慧优势,是指他在某个方面的智慧特征具有突出的表现。如果一个人以先天的智慧优势,来开发自己的智能,将能获得最佳的开发效果。每个人的智慧类型很少是单一型的,多数是两种以上智慧类型的混合型,只是以哪一种智慧类型为主罢了。我们进行智慧分型的目的,主要是为了便于研究和评价个人所具有的智慧特征与优势,为他们选择工作、事业发展提供一些有价值的参考。

从了解儿童的目的是为了帮助儿童发展的观念出发,要帮助儿童发展就要了解儿童。如何对儿童的多种智能做出正确的评估,描绘出每位儿童的智能分布呢?有研究认为,人类个体所拥有的各种基本智能或潜能,在儿童时期就已演化至能够展现其一种或数种优势的阶段。因此,观察儿童在

学习和生活过程中所展现出的某种优势潜能或倾向,对其每项基本智能遵照各项潜能的核心内涵和发展规律做出有前瞻性的评估,描绘出其智能结构的现有分布和可能的演化,显然是十分重要的。这就意味着,评价要尽量设法识别出儿童的智能强项和弱项。

但识别出儿童智能的强项和弱项并不是评价的最终目的。教育的主要功能应是在此基础上引导儿童利用优势领域的经验或潜能来进行学习,并以优势领域的智能来带动其智能的全面而和谐地发展;或在充分肯定其智能强项的基础上,针对其智能弱项采取有效的补救措施,扬长补短,达到促进其发展的目的。我们强调要识别出儿童的智能强项,是因为关注儿童的智能强项可使儿童获得成功的体验,感受到自己是有能力的,从而帮助他们树立自尊心和自信心,进而达到促进其全面而和谐地发展的目的。更重要的是,它可激发儿童的主体精神,为儿童的可持续发展奠定基础。

多种多样的活动,会给儿童提供一个展示自我、表现自我的平台和机会,使每个儿童的潜能得以展现。在此基础上通过观察儿童在学习和行为过程中所展现的某种优势潜能倾向或特定智能的不恰当行为,经综合分析后就有可能绘制出儿童的智能光谱—其各种基本智能强弱的剖面图。因此,基于多元智能理论的“多彩光谱”计划所提出的方案是:在教室设置大量能启发儿童运用多种智能(包括语言、数学、运动、音乐、科学、社会理解能力和视觉艺术等)的材料,并在每个领域都设计特别的游戏或活动,以更准确地确定儿童在该领域的智能程度。儿童在这样丰富的环境里学习和活动,就有充分的机会探索各个学习领域,展现各项潜能。在一个较长的时间内,充分观察和记录儿童在活动中的情况,就可以逐渐发现儿童的兴趣和才能,识别出儿童的智能强项和弱项。无论是家长或教师,要了解儿童,就应学会运用观察的手段,关注每一个儿童的智能倾向,包括观察儿童在各种活动中愿意做什么,不愿意做什么,观察儿童在各种非正规的活动中兴趣和活动及其交往水平,观察儿童与同伴交谈和讨论问题的情况等。通过长期的观察和了解,就可以了解儿童的个性特点,从而为进行有针对性的教育或为帮助儿童发展提供必要的基础。

### 5. 测评要尽量采用情景化的测度方法和实践检验的方法

情景化评估是根据智能的情境化提出的。由于智能在一定意义上应是一种在特定的文化和社会环境中解决问题和创新的能力,因此,要真实而准确地反映这种能力,就必须在实际生活或类似实际生活的情境中进行。情境化测度方法对成人来说就是尽量在实践中检验,在实践中来考察一个人的能力。而对儿童来说,情境化测度方法也意味着测度应该在真实的生活情景中或类似实际生活的情境中进行。要真实地反映儿童发展的状况,观测和评价就要贯穿于儿童日常生活的整个过程之中。这一方面是要求在实际生活情境中对儿童进行测度,另一方面,也是要求在儿童自然而真实的生活通过观察进行测评。当然,这就要求更多地发挥多个评价主体的作用。比如,家长要观察记录儿童在家庭和社区的活动的情况,并及时和教师交流和沟通;教师则应该从传统的测度方法中走出来,增强评价意识,在日常生活中随时随地地对儿童进行观察、记录和评价。儿童自身的反思和自我评价,对儿童自身的发展和评价有时也能起到一定的作用。只有通过这种生活化的测度,才能真实而准确地对儿童作出评价,才能通过评价促进儿童的发展。

更有人认为,情境化测评方法应该成为儿童学习情境的一部分,而不是在学习时间之外另行增加的内容。测评应该是在个体参与学习的情景中轻松地进行的。这就意味着,测评可以无处不在,若非特别需要,没必要去为测评而测评。

### 6. 测评要关注过程

关注过程,在智能的测评过程中,无论是对成人还是儿童,都是十分重要的。

关注过程,对成人来说,就意味着,我们在考察或评价一个人时,不只看其一时一事,还要看其整个经历和业绩;不只看其处理问题和事情的结果,还要看其处理问题和事情的整个过程。

关注过程,在对儿童进行测评时也很重要。从帮助儿童发展的目的出发,这就要求我们首先要建立起一套全过程测评的机制。要求测评要与实际的教学和训练融为一体,实现教学和测评的一体化。这也就意味着,教学的过程实际上也就是测评的过程,教学和测评是处于一个统一体之中。这种测评方法可使测评在一种无意识的状态下进行,成为自然的学习环节中的一个再平常不过的组成部分。其次,要将过程测评和结果测评结合起来,在评价儿童发展时不单凭某一结果去做判断。传统的对儿童进行测评的方法的严重弊端之一,就是只根据某一结果就对儿童做出评价,评价不能反映儿童实际的能力,不能反映儿童发展的过程。在对儿童进行测评的过程中,我们要更多地去关注过程,关注儿童在活动中的表现和成长进步等。

基于多元智能理论的评价观认为,教学和评价是“一体两面”的,评价要与实际的教学情境结合在一起。这样一种情境化的评价,使教师在教学过程中不会花太多的时间去进行测验,而只是把测验看作教学过程中的一个组成部分;学生也不再把考试看作可怕的“审判日”,评价贯穿于教学的整个活动之中,教学和评价的一体化就意味着教学活动的过程就是评价的过程。

传统的智力测评的严重弊端之一,就是评价与教学的分离,因而评价不能反映儿童实际解决问题的能力,不能反映儿童在学习活动中表现出来的情感态度和价值观等。而关注过程的评价则至少避免了这一弊端,把儿童在学习和活动过程中表现出来的兴趣、动机、努力程度都纳入评价之中,把儿童完成整个任务的过程做出完整的评价。这对儿童进行科学的、有效的评价和帮助其发展,无疑是非常有利的。

### 13.2.2 人类个体智能发展水平和个体差异测评的主要方法

#### 13.2.2.1 人类心理测评概述

心理测评是运用一系列心理学方法来测量和评价被试对象的心理状态或其他方面的个性差异的科学方法。是一种依据心理学原理和技术,以客观的、标准化了的程序对人的心理现象或行为进行量化的测量和确定,以判定个体差异的过程。

心理测评的基本问题是心理可否测评和如何测评。即心理的可测量性问题及科学测量问题。桑代克认为,“凡是客观存在的事物都有其数量”。麦柯尔则指出,“凡有其数量的事物都可以测量”。卡特尔则提出:心理学不根基于实验与测量上,决不能有自然科学的精确。他们都对心理的可测量性问题做出了肯定的答复。

心理的科学测评则比较复杂。心理测量与自然科学的测量不同。首先,是其测量工具的信度和效度不同。自然科学的测量可以重复多次相同的测量,而心理测量有时也可以重复,但多了会导致疲劳;有时还不能重复。另外,自然科学通常只测一个对象,即可推断总体;而心理测量常常需要测量一组对象,才可以推断总体,或推断个人与该组对象的关系。因此,心理测量远比自然科学的测量复杂。

对心理的测评包括测量和评价。其中,测量(measurement)是按照某种规则使用某种工具对某一特征赋值的过程。测量的核心要素包括法则、事物、数据。其中,法则就是测量的标准,或是测量的依据和准则,即我们依据什么来进行测量。法则有优劣之分。优良的法则可保证测量结果的准确性,所以法则通常有一个渐进的、不断完善的过程。事物即测量的对象,也就是我们要对什么进行测量,再明确一点就是哪些才是引起我们兴趣的事物的属性或特征。数据是测量结果的表现形式。

以数字表示测量结果常是区别测量与定性评价的一个标志。心理测量具有间接性、相对性、复杂性和针对性的特点。这里，间接性是说，心理测量是间接测量，其结论通常是从行为模式中推测出来的；相对性是说，每一个人被测得的结果都是与所在团体或人群的大多数人的行为或某种人为确定的标准相比较而言的；复杂性是说，测量的对象和测评的过程是复杂的；针对性是说，测量通常是针对一定目的进行的。评价（evaluation）也叫评估、评定。是对评估对象进行质或量的描述和价值判断的过程，是依据一定的标准，通过系统地搜集信息，在对标准与信息比较的基础上作出价值判断的过程，也是按照某种规则借助主观判断对某一特征赋值的过程。它可以在测量的基础上进行，也可以不用测量。心理评估（psychological measurement）则是应用心理学的理论和方法对人的心理品质及水平作出的鉴定的过程。心理评估的基本程序通常为：确定评估目的→详细了解被评估者当前心理状况→对重点问题深入了解和评估→资料分析、处理→结论。常用方法的评估方法则包括测试法、调查法、观察法（自然观察法和控制观察法）和会谈法（自由式会谈结构式会谈）等。

评估通常是依据某种价值观对事物及其属性进行判断和衡量。价值观不同，评价的结果必然不同，所以，价值观在评价中常起着核心的作用。哲学史上主要有三种不同的价值观：一是客观主义的价值观，认为价值是客观对象固有的本性，是纯客观的东西；二是主观主义的价值观，认为价值是用来表达个人对事物的好恶情感的，是纯主观的东西；三是辩证唯物主义的价值观，认为价值是一种表明客体对主体的特殊效用性，是主观性与客观性的统一。客观事物的效用性与主观的需要和愿望相结合，就具有一定的价值，符合的程度越大，价值就越大。显然，第三种观点是最符合实际的。

心理测量会有误差。心理测量的误差一方面来自测量工具、测量过程；另一方面也源于其间接性，即测量对象大部分并不能直接测量。

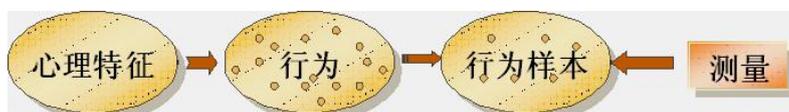


图13.2.1 人类心理测评的方法和过程

关于评价与测量的关系：测量是评价的依据，评价需要测量的具体支持。测量关心的是数据多少，评价关心的是价值如何。测量是一种纯客观的单一性活动过程，评价则是带有主观性和综合性的活动过程。心理测评的基础是心理测验。心理测验（psychological test）是根据客观的标准化程序对个体的某行为组样进行测量，以判定个别差异的过程。心理测验包括智力测验、能力测验、人格测验及诊断测验等。心理测验的要素包括：测量对象[心理行为]；测试方法[如，纸笔测试法；投射法—所谓投射法就是让被试者通过一定的媒介，建立起自己的想象世界，在无拘束的情景中，显露出其个性特征的一种个性测试方法；心理实验法—是指有目的严格控制或者创造一定的条件来引起个体某种心理活动的产生，以进行测量的一种方法]；常模[常模是测量中的比较标准，通过它可以确定测试结果的水平]；行为样组[行为领域的具有代表性的一组行为样本]；心理测试评价等。

心理测量一般要求应标准化，并最终实现客观化。要标准化就要求测验编制、实施、计分以及测验分数解释程序等都具有一致性。由于人类个体的行为常受环境的干扰而不一定真实反映心理，客观性就要求测验不应受主观支配，其测量方法应是可以重复的，测验的实施、计分和解释都应是客观的。评价心理测验的客观性的指标有：信度、效度、难度和区分度等。

信度[reliability]主要反映测试的可靠性、稳定性和一致性。亦即测量结果是否真实、客观的

反映实际水平。可以从以下三个方面理解测量的信度：① 实测值与真值相差的程度。也就是误差，误差越小，测量的信度越高。② 统计量与参数之间的接近程度。统计量指样本上的数字特征。参数是总体上的数字特征。统计量越接近参数，可靠性越高，信度越高。③ 两次重复测量或等值测量之间的关联程度。两次测量的相关系数越高，则测量的信度越高，但被测对象的经验、知识增长等因素也影响信度。信度包括再测信度[在不同的时间使用同一种测试工具或手段进行测试，两次结果之间的相关系数称再测信度—同一测验对同一组被试的先后两次测量之间的相关]；等同信度[复本信度][使用两种等值样本的测试工具进行测试得出的相关系数—两个内容、结构相同的测验对同一组被试测量结果的相关]；分半信度[将测试工具的内容分成两半，根据两半测试所得的分数计算出的相关系数，即分半系数]。目前，人们所确定的信度的可接受条件是：对一般能力与成就测验，信度系数应在0.90以上；对人格、兴趣测验，信度系数应在0.80-0.85之间。信度系数大于0.70时，可以用于团体间比较；信度系数大于0.85时，可以用于鉴别个人。

影响信度的因素较多。从测试工具方面讲，题意是否明确，测验的长度是否合适，测验分数的分布范围是否恰当，测验的难度是否适中，测验题目的客观化程度如何；从被试者方面讲，其心理是否健康，动机和态度如何；从主试方面讲，其专业水平如何；还有测试的环境状况；都会影响测量的信度。

在心理测量中，提高测量信度的方法有：① 适当增加测量题目的数量。数量多则代表性就大，既可提高信度，又可提高效率。但数量太多又操作困难。② 测验的难度要适中。难度适中，则信度大，区分度也大。③ 测验的内容应尽量同质。若测验内容庞杂，则增加难度，必然降低信度。④ 测验的程序应统一。测验的过程直接关系到信度的高低，信度需要规范公平的测验过程去保证。⑤ 测验的时间要充分。测验的时间如果不足，不能客观反映学生的实际水平，也影响信度。⑥ 评分要尽量客观、准确。评分不准确，必然降低信度。

**效度 (Validity)** 是指测量结果的准确性和有效性的程度，或一个测验对它所要测验的行为特征进行准确测量的程度，或测试绩效与实际工作绩效的相关程度。亦即测量是否达到了预期的目的，或者根据测验分数所作解释的有效性程度。可以从以下三个方面理解效度：① 效度始终是对一定的测量目的而言的。一般而言，判断效度的高低，就是判断测验达到目的的程度。② 效度也是对测量的结果而言的。测量结果真实、正确的反映被测对象的功能和特性时，则为效度较高或有效。③ 效度只是高与低的问题。测量的效度不存在有没有，只存在高与低。由于心理测量对象的特殊性（精神的、心理的、可以掩盖的），所以其测量更注重效度问题。

效度包括：① 内容效度：是指测验目的代表所欲测量的内容和引起预期反映所达到的程度，也就是测量内容的代表性程度。② 预测效度 (predictive validity)：也称效标 (criterion) 关联效度、经验效度或统计效度，是通过测试结果与实际工作绩效之间的联系来证明测试是否有效的一种效度类型。是以测验分数和效标之间的相关系数来表示测验效度的高低。效标就是鉴定效度的参照尺度。效标具有可变性，而且选择是比较困难的。③ 构想效度 (construct validity)：是指一个测量能实际测量出理论上的构念或心理特性的程度。即就是从心理学的理论观点对检测结果加以解释和探讨。

影响效度的因素包括：① 测验本身的因素（如，题目、结构、测试的长度、被试的选择）；② 实施；③ 信度。信度是效度的必要条件，不是充分条件。而提高效率的方法有：① 控制系统误差；② 精心编制量表；③ 妥善组织检测；④ 扩充样本的内容和代表性；⑤ 合理处理效度和信度的关系；

⑥ 适当增加测验的长度。

**难度**是指测验的难易程度。心理测量中,测题的难度一般用正确回答人数与参加测量总人数的比值表示。难度是测题对被测的适合程度的指标。难度既影响区分度,又影响信度和效度。但是难度是一个相对的概念。

难度会影响到测验分数的分布形态、测验分数的离散程度和测验的鉴别能力。测题的适宜难度,应该保证测验分数呈正态分布。影响测量难度的主要因素有:所测内容的多少;所考查能力的复杂程度或层次的高低;被测对测题的熟悉程度以及命题的技巧性等。

**区分度**是指测验对被测实际水平的区分程度。要想达到理性的测验信度,提高区分度是一个好方法。难度适中时,区分度有可能达到最大。提高区分度的方法是使测题和整个测试难度适中。

为了是测评更加可靠和有效,在进行心理测验中,还应该尽力做到:① 要按目的选择合适的心理测验;② 实施方法、过程和工具要尽量标准化;③ 对测试环境、主试、被试应有一定要求;如,要求测试环境要自然、协调,尽量排除外来干扰;要求主试应具有心理学专业的知识,具有心理评估与心理测量专业的知识与专业训练,具有善于捕捉被评估者言语和非言语信息的观察能力,具有自我认识和认识他人的能力,具有排除个人因素(情绪、价值观、道德标准、宗教信仰等)的影响对测试结果进行公平评估的能力,具有与被试进行沟通的能力;要求被试在测验时生理和心理状态良好,积极配合等;④ 能正确解释和对待测量结果。

心理测验的类型很多,从总的来看,心理测验主要包括(认知)能力测验和个性(心理素质)测验。认知能力测验包括一般能力测验和特殊能力测验。一般能力测验主要是智力测验,它主要提供对人的一般能力的了解,它并不能完全解决选拔与使用各种专业人才的需要。许多研究表明,智力和各种特殊能力之间的相关有时并不大,各种特殊能力都有自己的结构。为了测定从事某种专业活动的的能力,就要对这种活动进行分析研究,找出它所要求的心理特征,然后根据这些心理特征,列出测验项目,设计测验,以便测量特殊能力。当今国际上主要的特殊能力测验有:音乐能力测验、美术能力测验、数学能力测验、文书能力测验、机械能力测验、飞行能力测验等。能力测验既包括对实际能力[ability]的测验,也包括对潜在能力[能力倾向、性向(apptitude)]的测验。个性心理测验主要用于测量个人的性格、气质、动机、情绪、兴趣和态度等,价值观也属于人格范畴。

本书所关注的主要是智能,智能测量就是根据测量学原理和方法对智能现象及其属性进行数量化研究的过程。主要是对被测者内在精神属性的测量,包括智力水平、知识水平、能力水平、人格特征等。

### 13.2.2.2 人类个体智能测评的主要方法

可对心理素质和能力进行测评的方法有许多,比如:(1)行为评定法。行为评定法主要包括观察法、谈话法、作品分析法、个案法四种方法。① 观察法。观察法是在自然条件下通过观察一个人的行为、言语、表情、态度从而分析其心理的方法。采用此方法必须使被观察者处于自然情境中,保持心理活动的自然性和客观性,这样获得的资料才会真实;不论是长期观察还是短期观察,观察者都要做到有计划。② 谈话法。谈话法是通过与某人谈话从而了解其心理素质和性格特征等的方法。使用谈话法一定要事先确定谈话目的,要对谈话中的内容加以分析,要采取多种多样的谈话方式,要保持谈话气氛的融洽、和谐、温馨。谈话法在心理咨询中应用很广泛,它对了解人的个性和能力、搜集资料、确定解决问题的途径,具有重要意义。③ 作品分析法。作品分析法是通过对一个人的作品,如日记、命题作文、信札、传记、试卷以及劳动产品等的分析,来间接了解其心理特征

的方法。这种方法一般用来收集资料,对研究人的性格具有辅助性的意义。④ 个案法。个案法是通过收集一个人的家庭历史、社会关系、个人的成长史等多方面资料,来分析和了解其心理特征的方法。实际上,个案法就是观察法、谈话法、作品分析法的综合运用。(2) **自然实验法**。自然实验法是目前研究心理采用较多的方法,它是实验者根据研究的目的创设实验情境,主动引起被试的某种心理特征的表露,然后经分析、概括来确定其特征的方法。一位苏联心理学家曾用该方法设计了冬夜拾柴火的自然情境,以研究儿童在困难条件下的意志特征。实验是这样的:实验者把一部分干柴放到离宿舍不远的但需走一段夜路的山谷中,把一些湿柴放到离宿舍较远的但一路有灯光的储藏室中。要求学生定期在夜晚去捡柴火(不指定地点),实验者则藏在岔路口的小房内观察。结果发现,一部分学生勇敢而负责地到山谷中取干柴:有的学生边走边埋怨;还有部分学生怕黑,宁走远路去储藏室取湿柴。在这个实验中,实验者真实地了解了学生性格和意志特征的差异。自然实验法最大的特点是简便易行,获得的材料真实可靠。(3) **测验法**。测验法是用标准化测验测定心理特征的方法,主要包括自陈法和投射法。① **自陈法**。自陈法也称问卷法,一般是让被试按一定标准化程序和要求一次回答问卷中的大量问题,最后根据测验分数和常模来推知被试属于哪种类型。② **投射法**。投射法是利用某些材料(一般是意义模糊的刺激),要求被试对刺激材料进行解释,让他们在不知不觉中将其思想、态度、愿望和情感泄露出来,从而确定其心理特征。当然,运用上述方法来评定人的心理素质和能力都各有优缺点,因此,进行心理测评时应该综合运用多种方法,取各种方法之长,避各种方法之短,以便对人的心理特征作出合乎实际的评定。

本章我们关注的主要是**智能测评**。智能测评是指根据一定的价值观或目标,运用可行的科学手段,通过系统的搜集信息、分析解释,对个体智能状况[现状与发展的可能性]进行判断,从而为优化智能、教育或选拔决策提供依据的过程。传统测评偏重于鉴定和筛选,是为了区分或选拔适合的人员;现代测评价更强调反馈和矫正。测度的目的,或是为了论证智能差异的客观性,或是为了提供适合儿童的教育,或是为了挖掘个人智能发展的可能性,或是为个人或团体“有的放矢”的提供一些带有指导意义的建设性意见,帮助人们更全面地发挥自己的聪明才智。

我们认为,智能的测度应是以人类的各种实践活动为基础的。人类每个个体都具有其独特的智能结构,它不会因各人所从事工作的不同而时有时无,它在任何活动中都应是有所表现或起作用的。并且,人类具体的实践活动的项目、内容和方式是千变万化的,也正是这种千变万化才显示出了人类智能的重要作用。特定的智能差异测度方法,在对个体智能的构成和差异的测定中,应该是与智能差异测度的任务和-content密切相关的。因此,在智能差异的测度中,我们应该根据智能差异测度的任务和-content,选用合适的测度方法。只有选用了合适的测度方法,测定的结果才会是真实可靠的。

由于智能测度的方法可因具体测度对象和测度目的的不同而有所不同,因而,智能测度的方法应是多种多样的。若归纳起来,可有以下几种类型。

**1. 观察描述方法**。此法可适用于不同年龄、不同职业的各层次人员。根据具体测度对象和测度内容的不同,可采用不同的描述方法。如文字描述、操作描述、行为描述和活动描述等。对于儿童来说,采用行为描述和活动描述可能更为恰当,譬如,绘画描述,可根据不同年龄特点的儿童,给出不同水平的一幅画,让相同年龄组的儿童临摹并用文字表达画的内容和意境等。从临摹图与文字描述中,就可直观地反映出此儿童智能水平的某些方面,比如说,至少能反映出各个儿童的临摹能力、观察能力、欣赏能力以及文字表达能力等,同时,也有可能反映出各个儿童的这些能力究竟哪一方面更占优势等。让儿童在一个丰富多彩的环境里学习和活动,使其有充分的机会探索多个学习

和活动领域,在一个较长的时间内,充分观察和记录其在活动中的情况,就有可能发现被观察者的兴趣和才能,识别出其智能的强项和弱项。运用观察的手段,观察儿童在各种活动中愿意做什么,不愿意做什么,有兴趣的是什么;通过长期的观察和了解,就可以了解儿童的个性特点,包括其智能的倾向等。

对成人的面试和考察也应属于本类方法。譬如某单位需要一名高科技人才,就到相应单位(大学或科研机构)去采访(被采访者可包括各类工作人员,如:行政、业务、后勤等),根据采访后的结果可初步选择几个重点考察对象,对初步选择的对象进行一一面试,询问他们各人的经历、爱好、专长和业绩等,并请他们回答有针对性的问题,以对他们进行综合性的评估等。

**2. 定性与定量相结合的综合分析方法。**此类方法在智能结构的因素分析中曾被广泛使用,我们认为,它们同样也适用于对智能的测度。定性与定量相结合的综合分析方法有很多,比较典型的有因素分析方法、主观因素分析方法和模糊综合分析方法等。对不同年龄、不同职业的被测人员,可根据具体情况选用不同方法。因素分析方法或主观因素分析方法比较适合对智能结构的分析。

模糊综合分析方法是定性与定量相结合的综合分析方法中较为科学的分析方法,它特别适用于对不同人群进行智能差异的测度。在测度智能的构成因素及发展水平时,可对智能的构成因素及其发展水平做定量分析与定性分析。此法既能反映出智能构成具有综合性这一特点,又能分析出与某项具体实践有关的智能构成中起主导作用或占优势的智能要素。

模糊综合分析方法也是一种基于比较和分析的因素分析方法。通过比较和分析,我们可以大概了解从事某项具体活动所需要的主要智能构成或某人的主要智能构成。而后,再采用定量分析的方法,就可得出对某人完成这一具体活动的最恰当的评价,比如,“好”、“较好”、“一般”、“较差”、“很差”或“较好”、“良好”、“优秀”等。这些具有一定隶属度的评价虽然模糊,但在一定的测试人群范围内,其所反映出的对某人从事某项活动的适应性水平还是比较客观的。

**3. 标准化的量化测验方法。**过去人们主要是凭经验来对人的能力进行评估,二十世纪初,人们开始运用智力量表来测量人的智力。这种方法曾被广泛应用于对学前儿童或处于小学阶段的学生进行测评,开始是智商(IQ)测验或智力测验,后来人们又提出了情商(EQ)测验和创造力(CQ)测验等。

① 智力测验。最先运用智力测验量表进行智力测量的是比奈和西蒙,比奈-西蒙量表是第一个智力测验量表。推孟对比奈-西蒙量表进行修订,制定出斯坦福-比奈量表。推孟用智商表示人的智力水平,其智商公式为:智商=智力年龄/实际年龄\*100。美国学者韦克斯勒制定的智力测量表分年龄组,其智商计算公式为:智商=100+15(测值-均值)/标准差(SD),这种智商称之为离差智商,用以表示一个人的智力在同龄人中所处的相对地位。

② 特殊能力测验。由于智商不能简单的推断一个人的特殊能力,故有人制定了各种特殊能力测验量表测量人的特殊能力。

③ 创造能力测验。高智商是创造能力的必要条件,但不是充分条件,智力也不能直接推断创造能力水平,故人们制定创造能力测验量表,专门用于测量人的创造能力。

如今,智商(IQ)已是一个为大众所熟知的概念。当智力测验最早产生的时候,人们就是用智商来表征一个人的智力的高低程度的。但是,人们在那时所测量的智能主要是语言文字智能和数学逻辑智能。之所以如此,是由于经典的智力测验—基于比奈表的智力测验主要是为了能将真正心智有缺陷的儿童与由于其他原因而造成的学业不良儿童区分开来。因此,在编制智力测验表时,其测

验的内容主要是以学校里的学业内容为基础,所测量的也主要是儿童们的判断、理解和推理的能力。这就形成了传统IQ测验的特点,同时也导致了智力测验能部分地预测儿童在学校中的学业成绩,却不能准确地预测他们长大成人以后的整个人生的成就等问题。

在传统的智力测验中,智商被认为是一个人的整体智能或能力的反映。人们十分关注智商的测验和测量。有大量的研究都曾致力于制定各种不同的智商测验量表,认为用它们可以预测人们以后的发展。现在,这种智商测验的方法已受到广泛的质疑。由于智商测量方法不能检测智力的过程,IQ分数的高低经常受到测验样本与受测环境的影响,仅凭一张智商测验表很难标定一个人在各项活动或操作中所反映出来的智能构成,尤其是很难反映出一个人进一步发展的潜力;因此,基于单维的智能结构理论所建立的各种智力测评量表是很难对一个人的智能结构进行测度的。智商最严重的缺陷之一就是智商完全等同于了学业能力。因而其测出的智能是很不全面的。当然,由于从智商测验所获得的许多信息所反映的,主要是那种生活在特定社会环境与教育环境里的儿童所学习到的知识,在考察儿童对某方面知识的获取与掌握水平时,智商测验方法仍有它的实用性。

至于目前流行于市面且影响颇广的“情感智商(EQ)”,也曾被一些人所追捧,认为它对一个人的成功有重要影响。但是,由于其测试方法同样也是基于传统智力测验的程式,具有极强的机械性和片面性,因此,单纯的情商测验也很难反映一个人的真正的情商,特别是很难分析因智慧与情感相结合所带来的复杂性问题。当然,在这个经济全球化的年代,国际化与团队合作是未来的大趋势,因此,从小培养儿童成熟的情商(EQ),对于儿童未来的发展会有举足轻重的影响。但情商(EQ)的培养绝对不能采用教条式的课程,其测验也不应采用教条式的程式。

创造力(CQ)的培养和测验也是一个较新的提法。创新是一个永恒的话题。创造力无疑也是一个人的智力的最集中的体现。但如何对创造力(CQ)的进行测度和培养还有待研究。有研究认为,人的左、右脑分别掌管着不同的领域:左脑主司理论,掌管着认知、分析、语言、逻辑、推理……等思考性活动,是侧重于示范化及合理化的;而右脑则主掌感性,掌管着音乐、绘画、感情、想象力、创造力……等感受性活动,是充满着原创性与想象力的。培养或发展创造力就应该同时开发左脑和右脑。

**4. 专项单科或多科综合测试法。**此方法多用于学生学业智能差异的检测,当然,也适用于对成年人的智能差异的检测。比如,数学考试,能部分反映学生思维能力的某些方面;语文考试,至少能反映学生语言文字智能的多个方面。而汽车驾驶员、工人操作的技能测试,亦能反映操作能力的某些方面。

我们认为,测试不应是单纯的书面测试,也应包括实际操作和行为分析等。是为了测试动手能力还是为了测试动脑能力,是为了测试书本知识的掌握还是为了测试实践经验的多少,都要根据需要而定。

**5. 实践检验方法。**每个人的智能构成是不同的,其所表现出来的智能特征也就会不尽相同。有研究认为,这主要取决于两个方面:一是先天的遗传;二是后天的教育与环境的影响。一个人先天所获得的智能的基质,是决定他后天智能发展的基础;而一个人后天所经受的教育、经历、经验及环境的影响,则确定了他独有的智能特征。一个人的智能优势,是指他在某个方面的智能强项,认为他在此方面会有突出的表现或发展潜力。如果一个人以先天的智能优势,来主动开发自己的智能,将会获得最佳的开发效果。

由于每个人的智能构成很少是单一的,多数是多种智能的混合,对一个人的智能进行测度的目

的,在很大程度上是为了分析其智能构成在哪一方面更占优势,以为他们选择工作或今后事业的发展提供参考。而对一个人的智能构成或智能特征进行测度,我们主张最好的办法应是在实践的基础上进行。由于实践是检验认知真理性的最真实而重要手段,也应是检验一个人智能构成和水平的最真实而重要手段。一切可以在实践中进行的测度项目,我们主张都应在实践中进行。

从一定意义上讲,实验也是一种实践。实验的方法可适用于对一些智能内容的检测,如感觉、学习与记忆等。而有些活动,采用“模拟”或“模型”实验的方法也不错。不过,此两种方法多运用于对智能构成中的单项或数项因子的检测,要进行智能综合测试,仅靠实验或模拟的方法肯定不行。实验的方法中又有一些具体的方法,如观察法、结果验证法、实验室仪器检测法等,在通过思维过程或认知行为测度智能的构成或智能的差异,这些具体方法都可以根据需要采用。为了使评估更加客观,除了在实践中测度以外,还可结合调查法和追踪法等。即进行有针对性的调查和追踪。亦可先行调查再进行追踪。

除上述方法外,还有场研究法、生活史法等,不再一一列举。

这里应当说明的是:①上述方法都是一些基本方法,而这些方法既可同时使用又可单独使用。②这些方法大多还是对一个人当前智能水平的某些方面的大概性评估,而不是对一个人未来发展潜力的评估。由于一个人未来的发展是受多方面因素影响与作用的,用这些方法来评估未来发展潜力,只具有参考的价值。③由于每个人的智能都有其独特的表现方式、每一种智能成分又都有多种表现方式,我们很难找到一个适用于任何人的统一的评价标准来评价一个人的智能。④由于各种智能成分在个体的智能结构中都占有重要的位置,都处于同等重要的地位,它们在每一个个体身上都有自己独特的表现形式,因此,我们不能说具有不同智能结构的人谁更聪明,我们只能说他们各自在那个方面比较聪明,以及他们各自怎样聪明。换句话说,我们必须清醒地认识到,智能是多方面的,智能的表现形式是各不相同的,我们判断一个人聪明与否的标准当然也应该是多种多样的。

### 13.2.3 智力个体差异测评方法的发展

心理和智力的测评产生于社会的需要,是个体的差异引发了心理测量运动。高尔顿的遗传理论为心理测验奠定了最早的统计学基础;卡特尔的个别差异研究提出“心理学应建立于实验和测量上,心理测量应有一个普遍标准”,是对心理测量必要性的初步解释。比纳(A. Binet)是心理测量的鼻祖,他是发明智力测验常模量表的第一人。在智力差异测评领域,“19世纪80年代曾是高尔顿的10年,90年代曾是卡特尔的10年,20世纪头10年则是比奈的10年。”

从比纳编制世界上第一个智力测验量表起,智力测验已历经百年沧桑。在这百年中,智力测验已传播到了世界上几乎所有国家和诸多领域,以满足对不同人群的不同需要。目前,研究者已经开发了多种类型的智力测验,这些智力测验蜚声世界,在教育及其他社会领域产生了极大影响。然而,随着智力测验的广泛运用,它也越来越暴露出一些弊端,并因此而受到人们的批评。针对传统智力测验的弊端,近年来,研究者们正致力于改变传统智力测验使之呈现出新的面貌。传统的智力测验正在发生演变。这些演变包括:在测验依据上从考虑项目的鉴别性、区分性到增大其理论含量;在测验内容上从仅仅着眼于学业到测量更广泛内容;在测验形式上从静态的、回顾性的智力测验到动态测验;在测验的编制和实施上努力改善文化不公平现象;在测验的方法学基础上从经典测验理论发展到现代测量理论。增加了对元认知的测量;开发出了更有效预测创造力的智力测验;在有针对性地诊断、预测基础上提出了一些干预措施;能力倾向测验也取得了一些技术上的突破;对情感适应、人际关系、动机、兴趣和态度等人格测验进行了开发;等等。

### 1. 智力测验的依据：从考虑项目的鉴别性、区分性到增大理论含量

在智力测验的依据方面，一些研究者指责传统的智力测验并没有从有关智力的基本理论出发，不是心理测量学理论化的结果。确实，最早的智力测验是为了区分出需要接受特殊教育的儿童，因此在项目的设置上更多地是出于特殊需要的考虑。以后的智力测验大多沿袭了原有的传统，在编制时更多依据项目的鉴别性、区分性而非基本理论。因此，这样的测验虽然能够很好地评价智力的个别差异，却无法使我们更清楚地认识到智力的本质。

基于传统智力测验在理论性上的这种不足，新开发以及修订的智力测验有意识地加大了理论含量。比如，考夫曼首先改变了以往的智力测验非理论化的特点，以发展中的智力理论为依据，先后编制了儿童评估量表(Kaufman Assessment Battery for Children, K-ABC)和青少年-成人智力量表(Kaufman Adolescent & Adult Intelligence, KAIT)，其中K-ABC重点放在信息加工上，分别考察了系列加工、同时加工、心理加工组合和成就等级；而由一个晶体智力量表和一个流体智力量表构成的KAIT则试图整合卡特尔的流体和晶体智力理论以及其他理论家所提出的成人智力理论。考夫曼以理论框架来指导量表项目编制的努力值得肯定，但是他所依赖的理论是否适合于测验目的、是否能得到广泛的认同还有待于时间的检验。另外，最新修订的斯坦福-比纳量表(Stanford-Binet scale-IV, 简称SB-IV)，其15个子测验也按照智力结构的三层次等级模型组织了起来，在斯-比量表的理论构想中，G因素是最一般的能力，晶体智力、流体分析智力和短时记忆是三个二级水平因素，第三级水平也有三个因素，即对应于晶体智力的言语推理和数量推理以及对应于流体智力的抽象/视觉推理等。

有些研究者则尝试了另外一条道路，他们正试图建构一套完整、系统的智力理论，以便根据智力模型的理论构架来编制智力测验，弥补传统智力测验理论含量的不足。比如，戴斯等人在神经生理学基础上，整合了信息加工心理学和心理测量学理论，提出了智力的“计划-注意-同时性加工-继时性加工”模型(简称PASS理论模型)。依据PASS理论模型，他们在进一步实验研究的基础上，编制了一个标准化测验—戴斯-纳格利尔里的认知评估系统(简称CAS)，该测验面向5岁至17岁的个体，目的是测量与学习有关但又独立于教育的基本认知机能。全量表由12种任务类型构成四个分测验，每一分测验有三种任务，分别测量了计划、注意、同时性加工和继时性加工；测验中既包括言语测验也包含非言语测验，呈现的形式既有视觉通道也有听觉通道。经大样本实验发现，CAS能够非常有效地鉴别认知薄弱的儿童，并为未来的干预提供基础。因此，目前CAS已经开始在临床研究中使用。CAS有比较成熟的理论和实验基础，并且经过较严格的标准化，与传统上缺乏完整理论支持、针对性较差的量表相比，CAS更系统，评估的能力更广泛，也更具针对性，是对传统智力测验的很好补充，而戴斯等人的尝试也为今后智力测验的编制提供了有益的启示。

### 2. 智力测验的方法学基础：从经典测验理论到现代测量理论

在智力测验的方法学基础上，传统智力测验以经典测验理论(Classical Test Theory, 简称CTT, 又称真分数理论)为出发点，虽然大大增强了整个测验过程的客观化程度，促进了测验法的发展，但是从20世纪70年代起，CTT也受到了研究者的批评。其中最主要的问题就是：真分数以及误差无关的假设与实际情况不相符合；测验的各种指标如信度、效度、难度、区分度等依赖于具体的被测样本，即样本相关；所测量的心理特质依赖于具体使用的测验，即测验相关。

20世纪70年代以来，以项目反应理论为代表的现代测验理论逐渐成熟，由于这一理论具有解决实际问题的更大潜能，因此被研究者越来越多地应用于智力测验的实践当中，以克服传统智力测

验在方法学基础上的局限。项目反应理论(简称IRT,又称潜在特质理论)假设被试具有某种相对稳定的、支配其对相应测验作出反应并保持一致性的内在特征—潜在特质,研究的是被试这种潜在特质与其对项目反应之间的关系。IRT所测量的指标与被试的潜在特质相联系;项目的特征参数如项目辨别参数、难度参数、猜测参数用数学函数来确定;测量的信度和标准差通过项目的信息函数来推算;最后在项目水平上进行分析,并对项目的反应作出定性结论,因此有效地避免了传统测验理论的样本相关和测验相关,表现出了诸多优越性。另外,虽然IRT与经典测验理论的核心部分都是数学模型,但与经典测验理论的线性确定性模型相比,IRT采用的非线性概率模型能够更好地反映个体心理现象的复杂性;运用IRT还可以更好地识别被试在某些项目上的特殊反应、反常反应,从而成功地对被试提供诊断信息,而不仅仅是提供简单的等级排序或是否达到相应水平的信息,因而更具针对性和指导意义。

根据IRT和有关的假设,人们已经提出了许多双参数、三参数模型,并研制了一些用于项目分析的软件包。现在,随着项目反应理论的成熟和应用推广,项目分析理论的测验编制、误差估计、项目分析等技术已经越来越多地应用在智力测验项目的编制、选择以及评估当中,许多能力测验采用了拉希模型作为项目分析和个体评价的依据,一些测验如K-ABC和新版的SB-IV也提供了基于IRT的测验分数,以IRT为基础的项目功能差异(简称DIF)诊断技术也被用来考察智力测验题目的偏倚性等。同时,IRT也广泛地应用于智能水平的自适应测验编制以及题库的建设中。基于英国能力量表(British Ability Scale, BAS)修订和扩展的差异能力量表(Differential Ability Scales, DAS),就体现了IRT及其实践上的进展,该量表运用拉希模型对每一个项目的难度值进行校准,应用IRT来评定量表的内在一致性系数,并计算出对应分测验上每个可能分数的精确信息和测量误差,保证了量表的科学性。

此外,其他现代测量理论和技术如概化理论、标准参照测量、协方差结构分析、元分析、项目差异功能分析、计算机化测验等的应用,也使智力测验的编制、施测和评价焕然一新。

### 3. 测验的形式:从静态、回顾性的测验到动态测验

在智力测验的形式上,一些心理学家认为传统的智力测验是一种静态、回顾性的智力测验,虽然能够测量儿童已有的能力以及所达到的水平,但是不能反映儿童学习新知识的能力即学习的潜在能力,对智力发展的前瞻和预期能力不足,而且结果常常会受到被试的社会、经济、文化、教育等诸多因素影响,因此往往会低估一些儿童的智力水平。

由于传统智力测验更多指向儿童的过去—那些已经发展形成的能力,不能很好测量儿童的现在和未来—那些正在形成和即将发展的能力,因此,一些研究者提出了动态测验的思想,即不仅测量个体已经发展的能力,也要测量正在发展的能力。动态测验由维果斯基最近发展区的思想发展而来。所谓最近发展区,代表的是个体实际发展水平和潜在发展水平之间的距离,“是那些还没有成熟但处在成熟过程中的功能,它们可能明天就成熟但现在还处在胚胎阶段”。维果斯基认为心理学家在评估个体的能力时不仅要具有回顾性,还应具有前瞻性,实际发展水平体现的是已有的心理发展,而最近发展区则体现了预期的心理发展。动态测验就是试图评估个体智力潜能、测量最近发展区的一种测验。

**动态测验**常用的范式是将测验与干预结合起来,通常要求被试在测验中学习。其程序是:前测—教学(训练)—后测,被试在两次测验之间的获益代表了其学习潜能。与传统智力测验相比,动态测验更多关注学习发展过程,关注主试与被试之间的互动,关注在学习过程中由主试对被试的学习

能力作出评价。已有研究表明,动态测验能够有效地反映被试已有能力和潜能之间的差异。

由于不仅评价儿童现有水平,还着眼于儿童的发展潜能即最近发展区,寻求将潜能变为现实的干预手段,所以,动态测验已经越来越受到重视。例如,有研究发现,弗瑞斯坦的学习潜能评估测验(Learning Potential Assessment Device, LPAD)与传统的智力测验相比能够更好地预测被试的学习能力。而布朗等人则将实验程序中的干预和之后的迁移进一步量化和标准化,大大提高了动态测验的信度、效度,也拓展了其应用范围。

动态测验为我们提供了一个看待智力本质以及测量智力的新视角,也更丰富了原有的智力理论,是静态智力测验的重要补充。但是,和静态智力测验一样,动态测验也只是测量了智力的一个方面而已,并不能就此取代静态智力测验。要全面评价儿童智力,把两类测验结合起来应是一个可考虑的方向。另外,目前的动态测验在理论和操作上还比较模糊,标准化程度还不高,较多依赖于传统的智力测验,因此还有待于进一步改善。

#### 4. 智力测验的内容:从着眼于学业到测量更广泛内容

在智力测验的内容方面,许多研究者批评传统的智力测验仅仅着眼于学业,视野过于狭窄。传统智力测验测量的实质是其所测的主要是与学业有关的智力,其内部效度就是力求与学校教育的内容、方式一致,而外部效度则追求预测学业成功,表现出明显的学业取向。其智力的内涵被局限于一个比较狭隘的范围内,应当与现实相联系的部分却被严重忽视。鉴于传统智力测验在内容上的这种局限,一些研究者主张从更宽广的视角看待智力和智力测验,应测量更加广泛的内容。比如,斯滕伯格强调智力不仅仅与儿童、青少年的学业相联系,还关系到个体工作以及人生的成功,并由此提出了成功智能理论。

斯滕伯格所提出的成功智能包括三个方面,除了传统智力测验所测量的分析性智能,还包括创造性智能和实践性智能。在成功智能理论的基础上,斯滕伯格和他的同事经过大量的实验和教育实践,设计出了相应的测验,旨在评价个体的分析性、创造性和实践性智能。该测验由12个子测验组成,其中对每种智能的评价包括三个多选测验(内容分别为语言、数字和图形)和一个问答测验。斯滕伯格和他的同事通过研究特别指出,实践性智能和创造性智能正是传统智力测验所忽视的。比如,实践性智能主要体现于潜在知识(tacit knowledge)中(所谓潜在知识,指个体在一个环境中为了有效地工作而必须知道、但又没有明确地被教导过、无法用语言表达的知识),实验发现,潜在知识的测验分数与传统智力测验分数不相关,说明潜在知识与传统智力测量的内容不同。而有关研究表明,潜在知识比传统智力测验能够更好地预测个体在实际工作中的表现。因此,斯滕伯格认为传统的智力测验不论是在测量的对象、材料的类型还是测量的方式上都不够完善,已经到了拓展传统智力测验的时候了。

对传统智力测验拓展的另一方向源于毕生发展心理学的兴起。信息社会对人才的要求使人们逐渐重视对成人能力发展的研究,认识到开发真正适合于成人的智能测验的必要。研究发现,传统的用于儿童的智力测量类型并不适合于成人,其中重要的原因就是,用于儿童的智力评估标准多与学业成就相联系,而对成人的能力的评估,不应仅仅着眼于一个狭窄的主流文化中可利用的知识,还应关注职业成功、工作表现以及一生知识的获得等等。因此,成人的智力和能力标准,无论就其内容而言还是就其范围而论,都应与儿童大不相同。比如,埃克曼在其所提出的智力发展的加工-人格-兴趣-知识理论(简称PPIK)中就强调,在成人的智力评估中应当扩展知识的类型,特别是那些对工作和成人生活成功重要的知识。他认为,传统的评估关注的是知识的历史成分,它们大部分来自于

正规教育,但智力测验还应注重那些经过努力和相当长的一段时间积累起来的知识,这些知识是通过成人生活经验包括工作、爱好等获得的。为了更好地测量智力的教育/文化知识成分,埃克曼等人从大学水平考试中抽取了20个知识领域的主要内容,编制了知识量表,并以此对不同人群进行了研究,发现当这些知识作为智力评估的核心成分时,智力并没有随着年龄而下降。

另一些研究者在其对成人智能发展的论述中提出,所谓智能发展就是不断地整合那些特定领域中的新知识和日益增加的处理信息的专业化程序,这是整个成人生涯认知发展的核心,因此,应强调领域特异性知识(DSK, Domain-Specific Knowledge)的重要性。同时,一些对专家-新手研究、天才个体研究等的实验也支持了DSK对智力成绩的作用。这些成果提示我们,扩展成人智力和能力测验的内容是十分必要的。斯滕伯格等人为拓展传统智力测验的内容做了很大的努力,但其编制的量表要成为具有普遍意义的智力测验还有一段距离。埃克曼等人的知识量表的局限则在于,所测量的知识依然主要是通过学业训练而不是通过非学业经验获得的。

#### 13.2.4 人类个体智力差异的脑电生理学研究与相应检测

人类的智力行为,就其本质来说,是大脑的高级神经活动。因此,研究遗传等先天因素在智力活动中的作用,探究个体智力发展以及个体间智力差异的生物基础,并在此基础上建构新的智力理论,是很有必要的。

智力差异的脑电生理学研究,近年来获得较快发展。其中,EEG和ERP作为两种主要的电生理学技术,已被应用于智力差异探究之中。在智力水平与大脑激活程度存在负相关,以及其相关性受到任务难度、内容以及被试性别的影响等方面的研究,也已取得一定的进展;对以往个体智力差异脑电生理学研究中的ERP智力差异研究采用的任务略显简单、难以体现智力活动的本质等问题,正在做出改进;利用多种脑电技术结合研究个体智力差异的趋势,已日趋明显。

心理速度与检测时(inspection time, IT)实验等的提出,是对传统智力测试方法和理论的一种发展。出于文化公平性的考虑,研究者一直尝试运用生理测验代替智力的心理测验,IT实验范式就是在这一背景下提出的。它们最早是被用于研究知觉过程的,后来被用于揭示智力个体差异的生理基础。1976年,Nettelbeck和Lally首次用IT进行IQ研究,并发现了IT与IQ之间的显著相关关系,引起了心理学界的关注。

目前,检测时实验多是以视知觉理论为基础。在视知觉领域,研究者倾向于认为个体是以量子方式从环境中抽取信息的,每一个量子被称为一个知觉量。Vickers的“知觉决策的累积模型”(accumulator model of perceptual decision making)描述了视知觉早期阶段信息抽取的过程,以及当信息累积到足够量以做出准确的辨别判断的机制。

IT实验是以“准确率”(accuracy)为评价指标的,它是刺激呈现时间长度的函数。所以,IT实验所测的并不仅仅是反应所需的时间。IT实验相对于反应时实验的优势是,在刺激呈现后有掩蔽刺激,可以排除高水平的认知策略的影响。由于任务简单,它避免了反应时实验中的多因素混合影响的问题,也避免了速度-准确性权衡效应的出现。有观点认为,IT所反映的本质是在迫选条件、后掩蔽模式下的知觉阈限值。近年来,信息加工取向的IT研究已成为主流,越来越多的学者认为,IT反映了一种核心信息加工能力的早期阶段,是形成多种智力的基本元素。IT是IQ最可靠、最强的信息加工指标。这一指标体现为信息加工速度,而信息加工速度的神经基础则是神经传导速度(nerve conduction velocity, NCV),后者与智力的相关在0.26-0.48之间。

在检测技术方面,最早应用于智力研究的脑电技术是脑电图(electroencephalogram, EEG),随着计算机科学的进步和发展,研究人员开发出了事件相关电位(event related potential, ERP)技术。ERP应用于智力差异研究始于20世纪60年代,早期的ERP研究虽然技术上还不完善,但它的确反映出人对某些刺激物的反应存在着智力差异。最初,一些研究者尝试用这种技术取代IQ测验以解决文化公平的问题,由于实施困难,这种取向逐渐消失。随

后,研究者开始广泛地采用 EEG 和 ERP 等脑电技术来探索智力差异的生理机制和心理机制。20 世纪 70 年代以后,尽管已涌现出正电子发射扫描 (PET)、功能磁共振成像 (fMRI) 等现代脑电技术,但 EEG 和 ERP 作为非损伤性、高时间分辨率、费用相对低的脑电技术仍然是智力差异研究的重要方法。总体上,智力差异研究中采用脑电技术可以分成两种取向,一种是要求被试在一段较长时间内(如几分钟)解决复杂问题的同时记录其大脑活动,如 EEG、PET;另一种取向是在一个较窄的时间窗口(如一秒以内)内观察一些简单刺激引发的大脑活动,如 ERP。

### 1. 智力差异的 EEG 研究

EEG 是指利用头皮上的电极持续记录大脑头皮上的电位变化。它最初被应用于癫痫症的诊断和治疗,由于这些电位变化是由大脑内部亿万神经元活动引起的,包含了大脑皮层神经活动的丰富信息,因此心理学家用 EEG 来研究与大脑功能活动相关的认知功能。EEG 是研究智力差异非常重要的工具,研究者应用 EEG 对智力进行了多方面的研究,并取得了丰硕的成果。一般而言,对智力差异的 EEG 研究主要通过分析两类 EEG 指标来探索 IQ 与神经生理现象的关系,一类是功率(power)和振幅,另一类是神经网络的联结性,比如一致性、位相差、非线性动力学脑电复杂性测度等。近年来,研究者在已有的基础上又发展了一些新的方法,比如事件相关去同步化/同步化(ERD/ERS)等。

EEG 脑电信号根据频率的不同分成不同的频段,如  $\delta$  波 (<4 Hz)、 $\theta$  波 (4-7 Hz)、 $\alpha$  波 (8-12 Hz)、 $\beta$  波 (13-30 Hz) 和  $\gamma$  波 (30-70 Hz)。研究者大多对不同频段的 EEG 计算出功率或振幅,与 IQ 之间进行相关分析,并采用不同的脑电分析处理方法得到不同的脑电信息。目前对 EEG 脑电信号的处理主要是以功率谱分析为主,它通过计算不同频段的功率,把幅度随时间变化的脑波转换成脑电功率随频率变化的谱图。其中不同频段脑电波的功率可以分成绝对功率和相对功率,绝对功率是指这一频段脑电信号平均幅值的平方,相对功率是指每个绝对功率的值与整个脑电功率之和的比率。另外,对脑电信号也可以进行频谱分析,方法是计算出不同频段的脑电波在脑地形图中的分布位置,如  $\alpha$  波大多分布在枕叶皮层。大多智力差异的脑电研究中均发现, $\alpha$  波的绝对功率与 IQ 之间存在正相关。Jausovec 的研究表明高频  $\alpha$  波 (10-12 Hz) 的绝对功率比低频  $\alpha$  波 (8-9 Hz) 的绝对功率与 IQ 的正相关更显著;Jausovec 的研究还进一步发现  $\alpha$  和  $\beta$  波的电密度与 IQ 呈正相关。

尽管传统上研究者一般使用功率和振幅这两个变量进行研究,但是这种方法却不能提供任何关于大脑不同区域相互联系的信息。众所周知,大脑不同区域的信息整合在个体认知过程中的作用十分重要,一致性分析法则可以提供两导联之间的 EEG 信号的同步化信息。一致性是指某一频率范围内同一时间下的两个不同信号之间的相关程度,这个信号可以是位相角度或位相位移,可以主要用来考察不同脑区电极之间的耦合情况。研究者发现,在额叶区存在着低一致性与高智商的显著负相关,也就是说,高智商的人不同脑区的耦合较弱,这说明高智商者的脑区存在较强的功能分化,这符合神经效能假说(neural efficiency hypothesis)。这一假说认为,智力代表大脑的工作效率,这种效率来源于对与工作相关的脑区的调用以及不滥用与之无关的脑区。

(1) 不同认知任务下的智力差异的 EEG 研究。智力差异的 EEG 研究大多采用比较复杂的认知任务,如问题解决、记忆、推理、言语以及视空间任务等。要求被试完成认知任务的同时,记录被试脑电信号并进行分析。大量采用不同认知任务的 EEG 研究一致发现,不同认知任务下的大脑活动与智力之间存在着负相关。这一结果表明,智力高的被试在问题解决时不如智力一般的被试大脑活跃。心理学家一直试图用神经效能假说来解释这个问题,他们认为智力水平高的个体其神经系统的运行效率比智力水平一般的个体要高,这可能是由于智力水平高的被试基本上不使用与特定任务无关的大脑区域,而只是集中注意力使用与之密切相关的区域。这个理论假说还得到了采用其他方法(如 PET、fMRI 等)进行的一些研究的支持。

近年来,事件相关去同步化(event-related desynchronization, ERD)在智力的脑电研究中得到广泛使用。ERD 最早由 Pfurtscheller 和 Aranibar 提出,是一种测量大脑皮质激活程度的有效方法。其原理是当被试进行某认知任务时,大脑某一皮层区域活跃起来, $\alpha$  波段的脑电节律幅度减少。因此,研究者假设, $\alpha$  波段的 ERD 大小能够反映

出认知任务所涉及到的大脑皮质神经元兴奋程度的高低。大量个体智力差异的 ERD 研究发现,智力水平高的被试在进行不同类型认知任务时呈现更小的  $\alpha$  波段 ERD。例如,Grabner 等选用了数字广度任务、中央执行任务 (central executive task) 以及计数广度任务 (counting-span task),发现被试的  $\alpha$  波段的 ERD 大小与 IQ 之间呈负相关。由于 ERD 的大小表示大脑皮层激活程度,因此智力水平高的被试在进行认知活动时,其大脑皮层的激活程度比智力水平一般的被试低。但是,也有一些研究报告 IQ 与 ERD 呈正相关。对此,研究者进行了进一步研究,认为出现这种矛盾结果的原因之一可能是任务难度的不同。Neubauer 等人采用推理任务对被试的研究发现,如果没有指导被试采用特定的策略,被试在进行较难的认知任务时,没有发现 IQ 与 ERD 的相关;如果指导被试采用一定的策略解决问题,则发现 IQ 与 ERD 呈负相关。Doppelmayr 等人也发现任务难度可能影响 IQ 和大脑激活程度的相关,研究者给被试在计算机上呈现瑞典文推理测验图片,并根据被试的反应正确率区分出不同的任务难度,结果发现,智力水平高的被试在  $\theta$  波段表现出明显更强的激活,而在  $\alpha$  波段则发现任务难度与智力水平存在显著的交互作用,即在简单的任务中高智商的被试激活程度低,但是随着任务难度的增加,高智商被试的激活程度也不断增加,而智力水平一般被试的大脑激活程度没有变化。由此,研究者认为在任务难度较低时,智力水平高的被试可能采用比智力水平一般的被试更有效的解决策略,占用的注意资源少,因此大脑激活程度低,这也支持了神经效率假说;而随着任务难度的增加,智力水平高的被试需要集中更多的注意资源来完成任务,因此其大脑激活程度也不断提高。Neubauer 等人用 ERD 的方法对被试进行的研究还发现,智力与大脑激活程度的负相关可能受到任务的内容和性别差异的影响,不同性别的被试只有在他们完成出色的任务上才表现出智力与大脑激活的负相关,如女性在完成语言任务,男性在完成视觉空间任务上。Neubauer 等人运用创造性问题解决任务对智力与大脑激活程度的关系再次进行的研究发现,创造性的问题解决一般都伴随着较低程度的皮层唤醒。所有这些结果都表明,智力的个体差异可能造成了不同认知需求下大脑皮层反应的差别。

(2) **休息状态下的智力差异 EEG 研究。**在休息状态下智力差异的 EEG 研究得到的结论并不一致。比如,Lutzenberger 等人提出智力水平高的被试在休息状态下表现出更加复杂的 EEG 波形;Anokhin 等人的研究则说明,这种结果无法得到重复验证;Jausovec 发现高智商的被试在放松的状态下的确表现出更加活跃的大脑活动,然而这一结果却在其他的研究中没有得到验证。Jausovec 等考察了被试在休息状态下的智力水平与 EEG 的关系,发现男性被试在整个大脑区域其激活程度与韦氏成人智测分数之间呈负相关,且智力水平越高的男性大脑区域之间的耦合程度越低;而女性被试却与之相反,仅在额叶发现两者之间呈正相关,且智力水平越高的女性被试大脑区域之间的耦合程度越高。因此,就目前的研究成果看,个体在休息状态下的 EEG 与智力水平的关系,尚未取得一致。

研究表明,EEG 是评估大脑认知加工神经效率的一种有效方法,这是因为它不仅具有很高的时间分辨率,能同步反映出神经元细胞的激活状态,而且可以采用不同的脑电分析方法对脑电信号进行处理,从不同角度深入探索个体在认知过程中的大脑活动机制上的差异。

## 2. 智力差异的 ERP 研究

智力差异的 ERP 研究常采用听觉 Oddball 范式[怪球作业],即不同类型的刺激物以不同概率呈现,目标刺激物呈现概率低,标准刺激物呈现概率高。在研究中让不同 IQ 的被试完成相关的作业任务,同时记录脑电。另外,实验所用的任务还有检测时、Sternberg 记忆搜索任务 (Sternberg memory scan)、工作记忆任务等。众多研究表明,智力水平高的个体和智力水平一般的个体在 ERP 成分的潜伏期、波幅以及波形轨迹上存在着差异。

(1) **ERP 成分与智力差异。**一般情况下,Oddball 范式中的目标刺激物诱发出一个波峰值在 300ms 左右的 ERP 成分 P3,因此,很多学者通过研究 P3 来探讨智力差异和神经生理机制之间的关系。比如,Chalke 和 Ert 首次探索了 ERPs 和 IQ 之间的关系,发现某些成分的潜伏期与 IQ 之间存在着负相关。随后的大量研究证实了早期成分 P3 的潜伏期与 IQ 之间负相关的存在。例如,一项采用反应时任务的研究发现被试韦氏成人智力量表的成绩与各类型刺激物呈

现后 300ms 出现的 ERP 正走向波峰 (P300) 潜伏期之间的相关为-0.36; Polich 等人在一系列采用听觉 Oddball 范式的研究中发现 P300 的潜伏期与 IQ 之间的相关在-0.35 到-0.45 之间。梁福成在研究中也发现儿童的瑞文测验成绩与 P300 潜伏期的相关系数为-0.55, 达到了非常显著水平。与 IQ 相关的另一个 ERP 成分就是波幅。Polich 等人在听觉 Oddball 范式的研究中并未发现 P300 波幅与 IQ 之间的相关; Barrett 和 Eysenck 发现被试在刺激后 250-500ms 内的 ERP 平均波幅与 IQ 间存在负相关 (约-0.50)。梁福成使用 Oddball 模式测定儿童的 ERP, 发现瑞文测验成绩与 P300 的波幅呈正相关, 但不显著; 施建农等人的研究则发现, 智力超常组的 P3 (或 P600) 的潜伏期短于常态组, 而波幅高于常态组, 且差异达到显著水平。De Pascalis 等人所做的研究发现, 高智力的被试 P3b 成分的潜伏期更短, 波幅更大; 通过线性回归分析进一步提出, 额叶区的 P3b 成分的潜伏期和波幅可能是测量智力高低最有效的生理指标。

很多研究认为智力水平高的被试 P300 潜伏期短, 这说明其信息加工的速度要比智力水平一般的被试快。而信息加工速度可能是智力的一种基本成分, 它与大脑神经通路的传递速度有关, 智力水平高的被试神经传递速度相对较快, 因此信息加工速度快, 潜伏期比较短。这种观点在其他相关的研究中也得到证实, Vernon 和 Mori 以及 Reed 和 Jecsen 都报告神经传导速度 (NCV) 和智力间存在正相关, 相关系数在 0.26 到 0.48 之间。

信息加工速度理论与神经效能假说并不矛盾。神经效能假说认为, 智力水平高的个体其神经系统的运行效率比智力水平一般的个体要高。神经系统的运行效率包含神经通路的传递速度, 智力水平高的个体神经系统的运行效率高, 其神经通路的传递速度也快。因此, EEG 与 ERP 智力研究的这两种理论假设本质上是一致的。相比之下, ERP 成分的波幅与 IQ 的相关结论并不一致, 这可能与多数研究采用任务难度过低有关。一般来说, ERP 波幅代表着大脑加工深度, 它受到任务难度的影响比较大, 而智力的 ERP 研究采用的任务难度大多比较低, 因此很多研究难以发现波幅与 IQ 之间的显著相关。

除了对 P3 成分的研究之外, 研究还发现高 IQ 和低 IQ 组在完成检测任务时刺激出现后 140-200ms 内 ERP 波形 (N1-P2) 差异显著。与低 IQ 的被试相比, 高 IQ 的被试 N1-P2 上升曲线波形走势更陡, 而低 IQ 的被试上升曲线波形走势则比较平缓。同样, 在运用 Sternberg 记忆搜索任务进行的 ERP 研究中也发现, 在大约 165ms 时不同智力水平的被试 ERP 波形走势出现差异, 高 IQ 被试波形表现出一种延迟的负走向。尽管实验采用的任务不同, 但这些研究结果一致表明, 智力对信息加工过程的影响的确发生在 ERP 早成分的某一特定时间段。

(2) **ERP 波形轨迹与智力差异。** Rhodes 等人运用 “map-reading wheel” 方法研究 ERP 轨迹长度和 IQ 之间的关系, 这种方法后来被 Hendrickson 等人引入智力差异研究并将之命名为 “string length”, 它是指测量某段假定时程内的 ERP 轨迹的长度, 并计算此长度与智力水平之间的相关。Rhodes 等对 0-100ms、0-250ms 以及 100-250ms 这三个时程计算其轨迹长度与韦氏智力分数之间的相关, 结果仅在 100-250ms 内发现轨迹长度与 IQ 间呈显著正相关。后来的研究也得到了类似的结果, 例如, Stough 等人通过对 20 名被试的研究, 报告在 100-200ms 内 ERP 轨迹长度与 IQ 之间的相关最显著, 相关系数为 0.60。Hendrickson 提出了神经传递错误理论来解释这一结果, 他认为智力的差异并不是由于个体中枢神经系统传递速度的不同造成的, 而是反映了个体在问题解决时神经传递产生错误的比率不同。高智商的个体错误率低, 因而个体的脑电轨迹在每一个 trial 中比较稳定, 产生出更为详尽的诱发电位, 因而产生出更复杂的 ERP 轨迹。但是随后却有研究发现近于零的相关, 甚至负相关的现象。针对这些相矛盾的结果, 研究者尝试从不同角度来解释 ERP 轨迹长度与 IQ 的关系。Bates 提出轨迹长度与 IQ 间的关系依赖于不同任务对于注意资源的需求, 耗费较多注意资源的任务会产生负相关, 因为轨迹的复杂度应该代表大脑的工作效率, 而在其他的任务中, 轨迹的复杂度则表示认知能力。另外, 有研究者质疑这种测量 ERP 轨迹长度的方法本质上究竟在研究什么。一些研究者将它与 ERP 脑电复杂性联系在一起, 而 Haier 等人的研究发现轨迹长度依赖于波幅的大小。总之, 智力水平和 ERP 轨迹长度的相关在采用 string length 方法时更加显著, 但是与这种现象相应的心理机制至今依然无法确定。

(3) **ERP 成分与信息加工时间。**Woodcock 和 Mather 的信息加工能力的层次结构模型认为,人的信息加工能力是分层次的。其中,感觉检测能力(sensory detector capability)在最底层,上一层是联想加工能力(association processing capability),再上一层是知觉组织能力(perceptual organisation),最上层是关系教育能力(education of relations)。从底部到顶部,认知加工的复杂性逐步增加。这个结构中相邻层级之间的相关高于不相邻层级之间的相关。该模型得到许多研究证据的支持,包括 ERP 数据的支持。

根据层次结构模型,信息加工存在时间上的顺序关系,底层的认知加工最先进。猴子在视觉分类任务中的脑活动过程示意图,可以清楚地说明这一点。在猴的视觉分类任务中,平均反应时为 250-260ms。信息经丘脑的外侧膝状核(lateral geniculate nucleus, LGN)传导,从视网膜出发,到达初级视皮层 V1 区。随后,信息在 V2 和 V4 区继续加工,到达前/后颞下皮层区(PIT 和 AIT),此处有特异性地对某类刺激进行反应的神经元。信息再传递给前额叶皮层(PFC),此处有负责物体分类的视觉神经元,对视觉信息做出分类判断和反应决策。随后,信号通过前运动皮层(PMC)和初级运动皮层(MC)传到脊髓的运动神经元,再发出运动指令使手指肌肉做出运动。由于猴和人类在种族上的相似性,人类的简单视觉鉴别任务如 IT,应该也遵循类似的由后头部到前头部的信息加工过程。

对于心理能力的不同层次,也可以通过 ERP 的不同成分反映出来。研究发现,ERP 潜伏期反映了皮层的信息加工时间,早成分与低水平的心理成分相关,与视觉加工(Gv)、加工速度(Gs)和短时记忆(Gsm)呈高相关,在时间上主要对应于感觉检测等过程;而晚成分与高水平的心理成分相关,对应于知觉组织和关系整合等过程。

由于智力反映了个体在神经加工速度上的差异,也反映了心理能量上的差异。一些研究也开始关注 ERP 和心理能量间的关系,主要指标是 ERP 的波幅和潜伏期。在文献中,ERP 潜伏期与智力的相关平均为-0.30 左右。对于 ERP 波幅,临床研究发现,许多病理条件都会导致 ERP 波幅的降低。在智力和 ERP 波幅之间的关系上,研究发现,儿童的 ERP 早成分(100-250ms)表现出智力和波幅的正相关,而晚成分(250-500ms)表现出智力和波幅的负相关。

(4) **ERP 测量与检测时任务。**运用 IT 作业任务,研究发现刺激呈现后 140-200ms 的波(N1-P2 波)能够对智力有差异的被试进行区分。高 IQ 和低 IQ 组在刺激呈现后 140-200ms 之间出现了 ERP 波形上的显著差异,高 IQ 的被试,N1-P2 上升曲线波形更陡,而低 IQ 的被试,上升曲线的波形较为平缓。这种现象在额叶和颞-顶脑区表现得更为明显。

Burns 等人运用 5 项韦氏测验任务:句子记忆(Gsm)、划消(Gs)、视觉完形(Gv)、图形词汇(Gc)和分析-合成(Gf),测量两种视觉 ERP:黑白格棋盘式反转 ERP 和 IT 刺激判断 ERP。结果表明,IT 与划消测验(Gs)表现出 0.42 的显著相关,但是与其它任务没有表现出显著相关。这与以前的结果一致。在 ERP 成分上,顶叶的 P100 波潜伏期与句子记忆、划消、分析-合成和 IT 表现出显著相关,N1 波和句子记忆之间也表现出显著相关。该实验还发现潜伏期 70ms(N70)到 240ms 的波与认知能力测验相关,并与 IT 相关,说明在神经生理过程中的加工速度因素至少是智力的部分贡献因素。该实验把 ERP 研究的传统指标进一步扩展到了 N1、P1、P2 和 N70 指标,特别是 N1-P2 斜率(N1-P2 slope)正逐步成为智力 ERP 研究中的重要指标。

上述结果与通过其它智力作业任务得到的结果一致。人们对刺激的鉴别和分类一般从 165ms 开始,运用 Sternberg 记忆搜索任务也发现 IQ 对于 ERP 波幅的影响发生在 165ms 左右。同时,IQ 越高的被试在 165ms 处表现出更长的负向转折。在对于 IQ、IT 和 ERP 的研究中,Morris、Alcorn 和 Crinella、Yu 同样发现 IQ 影响 ERP 发生在 165ms 左右。这些证据都表明,智力对认知加工过程的影响发生在 ERP 早成分的一个特殊时段。

### 3. 人类个体智力差异的脑电生理学研究的进展

回顾 20 世纪心理学的发展,多数重大的发现都与一定的研究方法和技术进步有密切关系。EEG 和 ERP 具有高时间分辨率、无创伤性、费用相对低廉的优点,因此在智力差异的脑生理学研究中得到广泛的应用,并取得了丰硕的成果。但是,关于智力差异的 ERP 研究还占少数,智力差异的 EEG

研究依然蓬勃发展,这与其他领域的脑电研究趋势不太一致。究其原因,或许是 EEG 研究中脑电信号的处理方法比较多样化,能够从多个角度分析研究结果。而 ERP 尽管近年来发展迅速,但其实验指标比较稳定且略显单一。更重要的是,与其他脑电生理学技术相比,ERP 由于技术上的限制,其研究一般采用的是相对比较简单任务。而智力活动是一项十分复杂的认知活动,智力差异的 ERP 研究如果采用简单的认知任务,个体智力的生理差异很难完全反映出来。近年来,EEG 和 ERP 空间分辨率比较低的缺点使得智力差异的脑定位研究受到了局限。同时其他脑成像技术也具有其自身的缺点,如 fMRI、PET 由于时间分辨率低,无法及时反映大脑活动。为此,研究者开始倾向同时使用两种脑电技术探索智力差异的大脑机制。

从研究结果来看,作为心理速度或信息加工速度的评估指标,检测时、反应时、ERP 或神经传导速度都表现出与智力 G 因素的较高相关。这些指标是从非常简单的基本作业任务中获得的,不但可以解决文化公平性问题,也是评估智力差异的稳定指标。智力的心理计量学研究出现的结果间的冲突,有可能在认知心理学和生理心理学层面上得到解决。总之,新兴的脑电技术,已促进了智力认知神经科学取向的迅速发展,并对传统的智力研究和理论提出了挑战。

尽管关于智力的神经生物基础的研究取得了很大进展,但是由于智力活动本身的复杂性,使得寻找其生物学基础的工作开展起来必然具有一定的难度,也使得以往对智力的研究存在许多问题。当然,任何一种理论的提出和技术的进步发展都是一个长期而曲折的过程。EEG、ERP 与其他脑电技术的结合将有助于克服它们自身的局限性,逐渐成为人类智力活动脑机制研究的发展趋势之一。尽管认知神经科学对智力差异的研究依然处于探索阶段,尚未提出系统的智力理论,但其意义是十分深远的,并且再一次引起了研究者对于智力差异生理机制的关注,区别于过去关于智力先天遗传与后天教养在理论上的争论,认知神经科学体系中蓬勃发展的科学的脑电研究方法为研究者彻底解决这一问题提供了可能。

当前,智力差异研究的趋势是,在同一项智力研究中同时运用认知的与生物学的两种研究途径。也就是说,把心理计量学智力关联到认知或生理心理学变量上,以便考察两者之间的关联是否与信息加工的生物学因素有关。不过,获得心理计量学智力测验与简单的认知指标和生物参数的显著相关并不困难,困难的问题是这些简单指标的有效成分是什么,以及它们与智力差异的关联机制是什么。这应该是未来研究的重点。同时,心理速度只是智力能力的一个方面,要更好地理解智力和智力差异问题,需要多学科的交叉与融合。

### 13.3 儿童的智力发育与发展测评

#### 13.3.1 儿童智力测评的重要意义

我们为什么需要对儿童进行智力测评。原因很简单,就是为了更好的培养儿童。儿童是祖国的未来,人类的未来。儿童的培养教育,意义重大。但帮助儿童的前提是要了解儿童。要开发和培养儿童的智力,教育就必须符合儿童的智力结构特点;就必须符合或者贴近儿童的兴趣;就必须符合或者贴近儿童的未来职业选择。研究认为,每个儿童都是不同于他人的独立个体,每个儿童的先天素质和智力结构都是不同的,如果不了解儿童本身的素质和智力结构,只是一味按照一个外界的统一标准去培养所有的孩子,那么,孩子的个性特质就会被淹没在群体之中。如何评估孩子的素质和智力结构,如何描绘每位儿童的智能分布?进行必要的智力测试和评估,是了解儿童的智力结构和心理功能的最好途径之一。当然,对儿童进行智力测试和评估的作用,并不仅仅是要测得一个“智

商”，更重要的是要在测验过程中，结合对孩子行为的观察和认知方式的考察，运用对儿童心理发展和智力结构的理论来分析、解释测试结果，从而找出孩子的优势和弱点，对儿童进行有针对性的教育和训练。

心理学家经过大量的研究后认为，智力发育在儿童中基本呈正态分布。也就是说，智力高度发育(智力超常)与智力低下(智力发育落后)的儿童均占少数，绝大多数儿童属智力发育正常儿童。智力高度发育的儿童称超常儿童或早慧儿童。超常儿童是指智力发展显著地超过同年龄常态儿童的平均水平，或具有某方面突出发展的特殊才能，能创造性地完成某种或多种活动的儿童。我国古代称超常儿童为“神童”，西方国家称超常儿童为“天才儿童”。唐代诗人白居易 1 岁开始识字，5 岁开始做诗，9 岁时已精通声韵。莫扎特 5 岁时就开始作曲，11 岁时已能创作歌剧。德国数学家高斯 3 岁时就会心算，8 岁时就会解级数求和的问题，他在数学上的重大发现大部分是在 14-17 岁做出的。他们都是早慧儿童的典型。二十世纪初，人们曾以智商作为衡量天才儿童的指标。例如，美国心理学家推孟曾首先用智力测验来鉴别超常儿童，并把智商超过 140 的儿童定为天才儿童。20 世纪 50 年代以后，许多心理学家认为，仅用智力测验来鉴别天才儿童是有局限的、片面的，应该将多种指标结合起来评定天才儿童或早慧儿童。例如，坦南鲍姆(A. T. Tannebaum)提出，天才儿童是由五种因素交互作用而产生的。这五个因素是：一般能力；特殊能力；非智力因素；环境因素；机遇因素。认为超常儿童通常具有这样一些共同点：① 浓厚的认知兴趣，旺盛的求知欲；② 思维敏捷，理解力强，有独创性；③ 敏锐的感知力，良好的观察力；④ 注意力集中，记忆力强；⑤ 进取心强、自信、勤奋、有坚持性。超常儿童个性发展的速度比常态儿童快得多，个性发展水平也明显地高于常态儿童。从总体上说，超常儿童的社会适应性比较好，情绪比较稳定，意志坚强，喜欢并善于智力活动，动机效能高，特别是成就动机的水平较高。超常儿童是难得的好材料，我们对他们理应采取特殊的培养。

与智力超常儿童相对应的是低常儿童。低常儿童是指智力发展明显低于同龄正常发育儿童的平均水平，并有适应行为障碍的儿童。低常儿童又称智力落后儿童、弱智儿童等。当代心理学常根据下列三个指标来确定低常儿童：① 智商明显低下。一般认为，智商在 70 以下的儿童是低常儿童。② 社会适应不良。低常儿童对周围的自然环境和社会环境不能适应。例如，不能从事简单的劳动，生活不能自理，在学校里不能跟班学习等。③ 问题发生在早年。低常儿童不是某一种心理活动水平低下，而是整个心理活动各个方面的水平都低下。他们的心理活动具有如下特点：① 注意方面，智力低常儿童一般缺乏注意力，对周围事物不够关心，注意广度也狭窄；② 知觉方面，低常儿童知觉范围狭窄，速度缓慢，内容笼统而不够分化；③ 记忆方面，低常儿童对词和直观的材料识记都很差，再现中会发生大量的歪曲和错误，缺乏逻辑和意义的联系。记忆的保持也很差，视觉表象贫乏、缺乏分化并且不稳定；④ 言语方面，低常儿童的言语出现晚而且发展缓慢，意义含糊，词汇量小，缺乏连贯性；⑤ 思维方面，低常儿童的思维带有具体性，概括水平很低，在归纳、推理和概念化上都有困难，限制其对抽象材料的学习；⑥ 个性方面，低常儿童在个性上表现出沮丧、缺乏自信、对人有敌意，情绪紧张、压抑，常常以失败的心情来对待自己所做的工作，思想方法绝对化，等等。

一般认为，造成儿童智力发展落后的原因是多方面的，包括：① 遗传原因，如，父母染色体异常或有代谢性缺陷；② 孕期或出生时原因，如，母亲孕期患病、经放射线照射、情绪不佳或营养不良；胎位不正造成胎儿脑损伤或缺氧；产程过长造成新生儿缺氧；产程过短造成新生儿颅内毛细血管破裂出血而影响大脑的正常发育；③ 成长原因，如，脑膜炎或脑炎的后遗症，脑创伤、营养不良、

缺乏早期教育等等。智力发育落后是当今世界面临的一个重大的医学和社会问题。首先要通过优生优育防止低常儿童的出现,但也要有培育智力低常儿童的特殊教育方法。低常儿童是智力低下是客观现实,但我们也不应低估智力低常儿童发展的可能性。研究表明,智力低常儿童经过培养其智商水平可大幅提高,并且,大多数最终能适应社会生活。对于低常儿童进行特殊教育的关键是要尽早发现,然后根据不同的情况采取相应的教育措施,并且把病理诊断和治疗、心理治疗和教育措施紧密地结合起来,促进他们的智力发展,开发其智力潜能。

绝大多数儿童是智力常态儿童,我们的普通教育也主要是针对智力常态儿童。不过,研究认为,尽管都是智力常态儿童,但每个人的先天素质和智力结构是不同的。人都有自己的强项和弱项,人类在儿童阶段已经演化至能够展现数种优势智能,他们在解决问题或智力活动中已可显示出处理数种不同符号系统的潜能,因此,对儿童的智力结构进行测试和评估,发现儿童的优势潜能和优势领域,经综合分析后确定出儿童的“智能光谱”,按照儿童的不同的“智能光谱”进行不同的个性化教育,是开发儿童智慧的最佳方法和途径。

需要指出的是,一个人的智力发展的水平并非是由“智商”确定的,造成个体智力发展差异的原因也是很复杂的。可能有先天因素,也可能有后天因素。我们并不赞成单纯以智商来作为评定智力发展水平的标准。我们研究差异的目的,主要是在正视个体智力存在差异的基础上,针对能力的差异采取不同的教育措施,进行个别化的指导,充分发挥他们的各自特长,以促使他们的才能进一步发展。由此,我们也可总结出对儿童进行智力测评和智力结构分析的作用:① 及早发现“天才或资优”儿童或低常儿童;② 及早了解儿童的“智力结构”;③ 及早了解儿童的“发展潜能”;④ 解读被试过去的心智发展状况;⑤ 分析被试现在的智力现状;⑥ 指导和促进儿童和青少年的未来发展。

### 13.3.2 关于儿童智力测评的理论基础

智力测验是一种重要的心理测量技术,其目的是希望以此来对一个人的智力水平高低或是否聪明做出评估判断,有时候也用于测查病人的脑功能变化。智力测验研究的本意主要是希望找到通过测验来衡量人的智力水平高低的客观的科学方法。尽管智力测验的思想在古代学者的著作中就有所反映,但是用科学的方法把测验编制成量表来测量人的智力的努力则是从法国心理学家比奈(A. Binet)开始的。后来美国的心理学家韦克斯勒(D. Wechsler)编制了新的智力量表。

目前,关于智力测量的理论基础主要是关于智力的因素理论和与信息加工理论。它们认为,人的智力主要是一种一般性的综合认知能力,包括学习能力、适应能力和抽象思维能力等,是一种潜在的、非单一的能力,是知觉、分析和理解信息的复杂混合体。1904年,斯皮尔曼提出了智力的G因素的存在,这成为智力测验最初的理论基础。其后,关注智力测验的心理学家们通过深入研究,提出了一系列相关理论,它们也就成为了各类智力测验的理论基础。归结起来,这些理论主要包括:

#### 1. 智力的因素分析理论[智力的心理测量理论]

智力测验是想通过编制适宜的测验,对智力行为加以测量,然后依据对测验分数的统计分析作出推论,来说明一个人的智力水平或聪明程度。这就涉及一个基本的理论问题,即智力应包括哪些方面的因素,其结构应如何确定。为此,斯皮尔曼(C. Spearman)提出二因素理论,瑟斯顿(L. Thurstone)提出群因素理论,卡特尔(R. B. Cattell)提出流体智力与晶体智力,阜南(P. E. Vernon)提出智力的层次结构模型,吉尔福特提出智力的三维结构模型。我们统称其为智力的因素分析理论或智力的心理测量理论。

Spearman的二因素理论认为,人的智力由一般能力(G因素)和特殊能力(S因素)确定。G因素,即一般能力,主要指抽象思维能力、解决问题的能力、注意力、创造力;S因素,即特殊能力,主要指在特殊活动中需要的能力,比如口头表达能力等。瑟斯顿的群因素论认为,智力是多因素的,涉及到语言理解能力、言语流畅性、计算能力、推理能力、机械技能、空间知觉(感官和经验正确判断空间关系)、知觉速度(脑子反应问题的快慢)等7种基本的心理能力。卡特尔则认为,智力可分为两种:流体智力和晶体智力。流体智力主要指一般的学习和行为能力,它是一种潜在的智力,是与先天遗传关系非常密切的灵性,主要与人的神经生理的结构和功能有关,很少受后天教育因素的影响。晶体智力主要指已经获得的知识与技能,它是后天习得的,主要由后天教育和经验决定的。是在流体智力的基础上通过后天学习获得的,一定是通过学习获得的人们一般知识、教育水平关系密切。加德纳(Gardner)的多元智能理论则包容了更大范围的各种能力,认为人有多种基本智能,每个人在不同领域的智力发展水平是不同的。现在,将这些理论统一起来还是很困难的,因而,对智力测验的研究还需要继续深入下去。

## 2. 智力的信息加工理论

从信息加工能力的角度来理解智力和智能是目前人们比较一致的看法。为此,关于智力的信息加工理论研究也就格外引人注目。在智力的信息加工理论中,斯腾伯格(R. J. Sternberg)的智力三元理论认为,智力是个体对内外情景中的信息刺激进行加工处理的能力,主要由三个部分组成:(1)成分智力,指个体智力活动所必需的内在心理机制,这种机制主要由三种智力成分构成,即元成分、操作成分和知识获得成分。成分是一种基本的信息加工过程。其中,元成分(Metacomponents)用于计划、控制和决策的高级心理过程;操作成分(Performance components)表现为任务的执行过程,负责执行元成分的决策;知识获得成分(Knowledge-acquisition components)用于获取和保持新知识的过程,如,接受新刺激,做出判断和反应,对新的信息进行编码和存储等。(2)经验智力,主要指个体运用既有知识和经验处理新任务、新情境时的能力和信息加工过程自动化的能力。它既包括有效地应付从未见过的新事物的能力,也包括自动地应付熟悉的事情的能力。(3)情境智力,主要是指个体有目的的适应环境、选择环境和塑造环境的能力。这也就要求我们要在一定的社会文化情境中去理解智力。戴斯(J. P. Das)等人则把信息加工理论与认知研究中的因素分析方法相结合,并以鲁利亚的大脑三级功能区学说为理论基础,提出了智力活动中的信息加工的三级认知功能系统模型—PASS模型。认为,任何智力活动都要涉及到三个认知功能系统,即注意-唤醒系统、同时性加工和继时性加工系统、计划系统。同时性加工同步地整合刺激,继时性加工涉及的是将刺激整合为特定的系列,使各成分形成一种链状结构。三个系统之间动态联系,又相互影响。其中,注意系统是基础,加工系统处于中间层次,计划系统处于最高层次,它负责监督、管理、调节其它的心理过程。如:计划需要充分唤醒,注意力要充分集中;个体如何编码、加工信息会受计划功能的影响等。智力的信息加工理论在解释智力的内在机制方面无疑是比较深刻的。但如何根据智力的信息加工理论来进行智力测验则还需要深入探讨。

## 3. 智力的认知神经科学研究

认知神经科学研究的任务是阐明认知活动的脑机制,揭示人脑智慧的奥秘。认知神经科学,包括认知神经心理学、认知心理生理学、认知生理心理学、认知神经生物学和计算神经科学,这些新学科还未成熟,它们对智力的研究也处于探索阶段,尚未提出系统的智力理论。但是,我们却不能因此而忽视智力的认知神经科学研究对智力测验的指导作用。智力是大脑神经系统的系统功能,智

力的认知神经科学研究不但可以揭示智力的本质和产生机制，也能指导我们应如何来测定一个人的智力水平。

### 13.3.3 儿童心理与智力的标准化测评方法研究

按测验编制的规范性划分，智力测评可分为标准化测验和非标准化测验。关于智力的标准测试方法的研究，从1905年“Binet-Simon量表”在法国正式发表后，经过100多年的发展，智力测验的作用已经从甄别学习困难的儿童发展到了解儿童的智力特点，构建优势智力结构，进而进行有针对性的学习和训练的阶段。在一些西方国家，每个人一生中至少要进行一次测力测验和智力结构分析，很多家长会在孩子幼儿阶段或少年阶段为其做智力测验或智力结构分析，从而发现孩子的智力潜能，预测其未来的发展方向，同时给予科学的训练和指导等。当前，国际上常用的智力测验有两种：斯坦福-比奈量表和韦克斯勒智力测验。

#### 12.3.3.1 关于智力标准化测验编制中的一些问题

对智力进行标准化的测试首先需要制定标准化的测验方法。一个科学测验在编制时要经过标准化的过程。所谓标准化一般包括：选择有代表性的测验题目—选取具有代表性的被试—施测过程标准化—建立常模等步骤和过程。心理测验编制的一般程序是：（1）测量目标分析。包括理论界定、结构分析、因素分析等。（2）编题。要求：形式上一般以客观题为主，编排上以让被试容易理解为主。（3）试测。通过抽样测试和定量分析，检测试题的可用性。（4）形成正式题本，进行信度分析、效度分析和项目分析。项目分析要对各项目的难度、反应度、辨别力进行分析。（5）完成常模和标准制定。常模[norm]是根据标准化样本的测试获得的具有代表性的结果。（6）编写测验手册。智力测验属心理测验，其标准化也应如此。

智力测验的标准化通常要求要努力克服文化不公平现象。在智力测验的编制和实施上，许多研究者指责传统的智力测验存在着文化不公平现象，即研究者所编制的智力测验主要测量的是他们所认同的适合主流文化的技能和知识，文化偏差造成了对其他文化的个体智力低估的现象。在智力测验的施测中也存在着这样的问题，由于施测者与被测者的文化背景不同可能导致的彼此不能很好沟通、反馈和相互领会，也会影响被测者的发挥及其IQ分数。

文化不公平现象是从智力测验产生伊始就存在的问题，至今仍然是智力测验领域的一道难题。为了克服这个问题，研究者曾经编制了受文化经验影响较小的智力测验，例如瑞文图形推理测验、卡特尔文化公平测验等。由于受心理测量学智力理论的影响，传统的智力测验曾一度倾向于测量所谓不受任何后天社会文化因素影响的“纯”智力，早期的认知心理学家也设计了实验任务，来测量在他们看来反映智力本质的所谓“纯”信息加工速度。在这些秉持纯粹智力观的研究者看来，智力应与社会文化因素无关，与具体操作情境无关，因此，他们追求的智力测验也是“纯粹的”、“脱离社会文化背景的”、“不受社会文化因素影响”的“文化公平测验”。然而，随着对智力研究的深入，人们已逐渐认识到，脱离了智力形成和发展的社会文化环境，孤立地研究智力，只能得出片面和狭隘的认识。特别是近年来随着心理学研究领域里生态化趋势的渐盛，随着社会文化理论和人类学取向的兴起以及心理学研究本土化的发展，人们又重新关注社会文化背景的作用，越来越认识到社会文化对智力、智力发展以及智力特征的制约性，进而提出了文化相对性的智力理论。例如，在斯滕伯格的三元智能理论中，其背景子理论就强调了智力是适应、塑造、选择环境背景所需要的心理活动，必须从其发生的背景中看待智力，因为在一个文化中重要的智力技能，在另一个文化中或许就没有那么重要，在不同文化中传承的智力技能也可能不尽相同。因此，斯滕伯格强调，智力测验中那些适应于不同文化情境的候选设计十分必要。因此，现代的智力测验已不再强调独立于人们所生长和发展的文化环境，而是试图从特殊的文化或者子文化的视角来进行。同样，也并不主张一味去寻找跨文化的有效测验，认为更重要的是去理解和预测不同文化情境下人们的行为，认为只有这样才能编制出适用于特殊文化的智力测验。可见，在改善文化不公平的问题上，研究者曾经走过弯路。目前，主要的问题已是如何结合人们成长和生活的社会文

化环境来进行智力测验的编制和施测的问题了。

智力测验的标准化也要求要努力克服智力发展的性别差异。男女两性智力的发展不是同步的，而且男女两性的智力可能存在着各自的优势领域（如，女性擅长言语表达、形象记忆等，男性则长于逻辑思维、空间推理等）。而且，从IQ分数的性别分布上来看，男性智力的变异要大于女性，在智力分布的两端（即特别聪明和特别愚笨），男性均多于女性。许多智力测验在编制和修订时，都会尽量删去那些存在性别差异的项目，但实际上，保留下来的项目未必就具有性别上的公平性。比如，韦克斯勒成人智力量表，对男性有利的有5个分量表（常识、理解、算术、填图、积木图案），而对女性有利的却只有3个分量表（类同、词汇、物体拼组）。有人认为，即使男女两性真的在智力测验上表现出差异，也可能是由于非智力因素所引起的。比如，在小学阶段，男孩和女孩在数学上旗鼓相当，但到了高中，男性将略有胜出，到了大学则占明显优势。汉森（Hansen）等认为，这一差异很可能是由男性和女性不同的归因方式而导致的。

弗林效应也是智力测验需要考虑到的一个问题。弗林收集了历史上大量的智力测验数据资料，经过系统研究后发现，自智力测验发端以来，智力测验平均成绩在不断上升。这种现象被称为弗林效应（the Flynn Effect）。弗林的结论是，从1940年开始，智力测验的平均成绩（IQ的平均分数）在以每10年3个百分点的速度递增，这意味着现在智力测验测得的100同20年前测得的106是相等的。而且群体智力测验平均分数上升的速度有加快之势。

衡量智力的尺度现在主要是智商—智力商数。智商（intelligence quotient）通常简称IQ，是个体智力水平的数量化，用以衡量智力水平的高低。比奈在量表编制中首先提出了“心理年龄[智龄]”的概念，用来表示儿童智力水平的高低，这一概念后来被发展为智商的概念。简单说，其智商是通过将心理年龄[智龄]（MA）和实际年龄（CA）之比乘以100而得到的。其计算智商的公式为： $IQ=MA/CA \times 100$ ，以此计算出来的智商叫比率智商。韦克斯勒

（D. Wechsler）又提出离差智商的概念。所谓离差智商是用一个人在他的同龄人中的相对位置来衡量他的智力水平，其计算公式为： $IQ=100+15Z$ ，其中， $Z=(\text{测值}-\text{均值})/\text{标准差}(\text{SD})$ ，离差智商反映的是一个人的智力在同龄组中的相对位置，因而不受个体年龄增长的影响。目前，国际上已普遍采用这种智商概念。

智力测验的标准化要求有合适的信度与效度。信度（reliability）即测验的可靠性，是指测量结果的稳定性程度。信度的大小通常用两个测量结果的相关系数来表示，称为信度系数。信度太低的测验不能使用。智力测验的信度系数一般要达到0.90以上。智力测验的效度（Validity）即测验的有效性，是指一个测验或量表实际能测出其所要测的心理特质的程度。效度通常用效度系数来表示。智力测验的效度系数一般在0.3~0.6之间。如斯坦福-比纳量表测得的智商与学生的学业成绩之间，相关可达0.6。

我们认为，智力测验是必要的，但用智商测量法来测量智力却是不可取的，因为智商测量并不具备测量一个人智力分布的功能，不能检测智力发展的潜力和智力心理过程。而现在所谓的“智商”，也不具备验证一个人实际智力发展水平的效能。IQ分数的高低，常常受到测验样本和常模的左右，受到被测者的家庭和教育背景的制约，也受被测者练习和经验的影响。有人曾提出基于人体结构的“物理测智法”以及基于人体生态的“生理测智法”等，也只是一种有益的探索。我们反对用一个人作为确定的统一模式来“标签”活生生的人类个体。标准化是需要的，但是，任何机械和片面的标准都难以反映人类的智能。科学的方法应是从发展的和辩证的观点来看问题，特别是对与智慧和情感相关联的问题，更是如此。可惜，要做到这一点，现在还有一定难度。

### 13.3.3.2 典型的标准化智力测验

目前使用的标准化智力测验，主要有：斯坦福-比奈智力测验；韦克斯勒（Wechsler）智力测验；瑞文（Raven）推理测验等。

#### 1. 斯坦福-比奈智力测验与智力测验量表

为鉴定儿童智力发育的需要，特别是为了鉴别智力低下儿童，1905年，法国心理学家比奈和医

生西蒙合作,在G因素理论的影响下,编制了世界上第一个智力量表——比奈-西蒙智力量表(B-S量表)。这套测验由30个从易到难的题目组成,可以测量智力的多方面表现,但主要测量的是人的判断力、理解能力和推论能力。1908年,量表进行了修订。修订后量表的测验项目增加至59个;适用年龄从3岁到13岁。该量表首次启用了智力年龄(mental age)这一概念,是第一个年龄量表。它按年龄对测验项目进行分组,每个年龄组设计了6个测题,每通过1个测题代表2个月的智力年龄,并以智力年龄(mental age, MA)来表达儿童的智力发展水平。1911年发表了比奈和西蒙的第三个智力量表。它与1908年的量表相比变化不大,只是在每个年龄组的测验项目上略有增删,并增设了一个成人组。

美国斯坦福大学的心理学教授推孟(Terman)将比奈-西蒙量表介绍到美国,并对比奈-西蒙量表进行了修订,使它进一步标准化,被称为斯坦福-比奈(S-B)量表[Stanford-Binet Intelligence Scale]。该量表1916年出版,共有90个项目。斯-比量表曾于1937年、1960年、1972年、1982年以及1986年几次做了修订,适用的范围自2岁至14岁,还另外设计了普通成人组和优秀成人组测题。成为当今世界上广为流传、极具影响力的智力标准测验之一。

斯坦福-比奈量表按照年龄组,每个年龄组6个项目,每道题目代表2个月。随着年龄组的上升,项目的难度逐渐增加。智力年龄也是斯坦福-比奈量表表达智力水平的一个重要概念。在使用量表进行测验时,如果一个10岁的儿童通过了10岁组的全部项目,那么他的智力年龄(简称智龄)就是10岁。如果他还通过了11岁组的两个项目(代表4个月)和12岁组的一个项目(代表2个月),他的智龄就是10岁6个月。为了了解一个儿童的智力发展水平,还必须考虑智力年龄与实足年龄的关系,为此,推孟提出了智商(IQ)的概念,作为比较智力水平的相对指标。智龄决定于个体所通过的项目所属的年龄。这种比率智商可用来对不同年龄儿童的聪明程度进行比较,但是,这种计算智商的方法有一个明显的缺陷,就是若一个人的智力没有变化,那他的智商会逐步减小,因为它是以智力年龄与实足年龄一起增长为前提的。而实际情况并非如此。人到了一定年龄,智力有可能不再随着实足年龄增长,而且还有降低的趋势。因此,若按传统的比率智商计算,一个人在20岁时智商为130,到了40岁智商则可能为65,成为低常了。为此,韦克斯勒的离差智商便应运而生了。

## 2. 韦克斯勒智力测验与智力测验量表

韦氏智力量表(Wechsler Intelligence Scale)由美国心理学家韦克斯勒所编制,是继比奈-西蒙量表(B-S量表)之后为国际通用的另一套智力量表。韦克斯勒为编制智力量表做出了巨大贡献。他首创离差智商(deviation IQ),用以代替比率智商,克服了比率智商的不足。韦克斯勒的离差智商实质上就是一个人的智力测验成绩和同年龄组被试的平均成绩比较而得出的相对分数,以确定被试的智力水平在同龄人中的相对位置。即一个人的智力偏离本年龄组平均水平的方向和程度。韦氏认为,一个人的智力水平可以用他的测验分数与同一年龄组其他人的测验分数相比较来表达,即采用离差智商作为估计智力的相对聪明的程度。某一被试的离差智商是视此人的测验分数与其同龄被试测验分数分布曲线上的均数的离差大小而定的。即被试的分数处于同龄标准化样组均数之上或之下有多远,离差大且为正数,则智商高;而离差大且为负数,则智商低。这种以个体在同龄群体中的相对位置来代表此人的智力水平,既简洁、明了,又客观、科学。韦氏将测试分数进行了标准化,其离差智商是根据团体的标准差和平均数来测量智商的,它表示个体的智力水平在其年龄组中团体中的相对位置。而这样做的一个基本假设是:在同龄人的团体测试中,人的智力测验分数将呈正态分布,若令所有同龄人的智力测验分数的平均水平(均值)为100( $IQ=100$ ),智商分布的标

准差为 15，则一个人的智力水平可以用一个人测示验分数与同一年龄组其他个体的测验分数相比来表达。韦氏量表测量分数的具体计算方法是：以 IQ=100 分作为中等水平，以 IQ 分正负偏离 100 分表达低智力和高智力。例如，某个年龄组的平均分数为 80 分，标准差是 10 分，某人得分为 90 分，他的标准分数为  $SX(90-80)/SX[10]=+1$ ，代入公式，则其离差智商= $100+15(+1)=115$ 。

韦克斯勒智力量表主要包括三个量表—韦氏成人智力量表(WAIS) [16-74岁]；韦氏儿童智力量表(WISC) [6-16岁]；韦氏学前儿童智力量表(WPPSI) [4-6.5岁]；各自独立又相互连接，适用的年龄范围从幼年到老年，适用的年龄范围比斯坦福-比奈量表更广。

韦氏从整体智力观点出发，把智力分成言语智力和操作智力两部分，相应由两个子量表—语言量表和操作量表—构成。他编制的各种智力量表都包括言语和操作两个分量表，可测量较为广泛的能力，得到言语智商 (VIQ)、操作智商 (PIQ) 和全量表智商 (FIQ)。言语量表和操作量表分别又有分测验，每个包括 6 个子测试项目。比如韦克斯勒儿童智力量表 (WISC) 中，言语量表包括常识、类同、算术、词汇、理解、数字广度六个分测验；操作量表包括图画补缺、图片排列、积木图案、物体拼组、译码、迷津六个分测验。除了可以计算全量表的智商外，还可分别计算言语智商和操作智商。

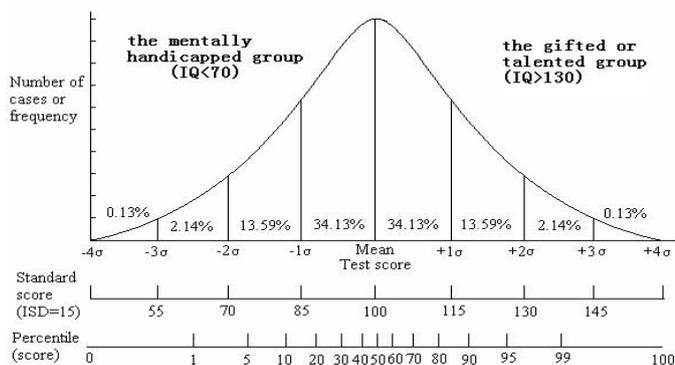


图13.3.1 韦氏智力测量分数正态分布曲线

韦氏智力测验如今是世界上使用最为普遍、最具权威的智力测量工具。共11个分测验，可以了解儿童或成人的以下方面：(1) 知识的保持和广度；(2) 实际知识的理解及判断力；(3) 心算及推理能力；(4) 抽象概括能力；(5) 注意力和暂时记忆力；(6) 词语的知识广度；(7) 学习和书写速度；(8) 视觉记忆和视觉-空间理解力；(9) 视觉-视觉空间的分析综合能力；(10) 对故事情境的理解能力；(11) 处理局部和整体关系的能力。采用韦氏智力量表，对儿童和青少年进行智力测评和分析，发现其智力状态和智力结构的特点，可为其进行个性化的智力发展指导和设计专项训练，使之在未来的智力发展中扬长而避短，更大限度地发挥其内在潜力。

韦氏智力量表各分测验的主要测试项目和功能包括：(1) 常识：用于知识测验。韦克斯勒认为，智商越高的人，兴趣越广泛，好奇心越强，所以获得的知识就越多。故此测验主要测量人的知识广度、一般的学习及接受能力、对材料的记忆及对日常事物的认识能力。测验由 29 个常识问题构成，包括历史、天文、地理、文学和自然等内容。(2) 领悟：此测验主要测量判断能力、运用实际知识解决新问题的能力以及一般知识。与知识测验相比，受文化教育影响较小，但记分难以掌握。(3) 算术：此测验主要测量数学计算的推理能力及主动注意的能力。该能力随年龄而发展，故能考察智力的发展，同时对预测一个人未来心智能力很有价值。(4) 相似性：此测验设计用来测量逻辑思维

能力、抽象思维能力与概括能力，是G因素的很好测量指标。(5) 数字广度：此测验主要用于记忆测验，测量人的注意力和短时记忆能力。临床研究表明，数字广度测验对智力较低者测的是短时记忆能力，但对智力较高者实际测量的是注意力，且得分未必会高。数字最长是12位数字。(6) 词汇：本测验主要测量人的言语理解能力，与抽象概括能力有关，同时能在一定程度上了解其知识范围和文化背景。研究表明，它是测量智力G因素的最佳指标，可靠性很高。但其记分较麻烦，评分标准难掌握，实施时间也较长。(7) 数字符号：该测验主要测量一般的学习能力、知觉辨别能力及灵活性，以及动机强度等。该测验与工种、性别、性格和个人缺陷有关，不能很好地测量智力的G因素，但具有记分快、不受文化影响的特点。(8) 图画填充：此测验主要测量人的视觉辨认能力，以及视觉记忆与视觉理解能力。填图测验有趣味性[每幅图少画一样，如没有尾巴的猪]，能测量智力的G因素，但它易受个人经验、性别、生长环境的影响。(9) 积木图案：该测验主要测量辨认空间关系的能力、视觉结构的分析和综合能力，以及视觉-运动协调能力等。在临床上，该测验对于诊断知觉障碍、注意障碍、老年衰退具有很高的效度。(10) 图片排列：此测验主要测量被试者的分析综合能力、观察因果关系的能力、社会计划性、预期力和幽默感等。它也可以测量智力的G因素，可作为跨文化的测验。但此测验易受视觉敏锐性的影响。(11) 图形[物体]拼凑：此测验主要测量处理局部与整体关系的能力、概括思维能力、知觉组织能力以及辨别能力。在临床上，此测验可了解被试的知觉类型，他对尝试错误方法所依赖的程度，以及对错误反应的应对方法。此测验与其他分测验相关较低，并对被试的鉴别力不甚高。在上述项目中，其中的数字广度[背数]和图形[物体]拼凑两个分测验是备用测验，当某个分测验由于某种原因不能施测时，可以用之替代。

测验实施时，言语分测验和操作分测验交替进行，以维持被试的兴趣，避免疲劳和厌倦。完成整个测验需约50-70分钟。所有分测验的原始分数都要转化为标准分数。分别将5个言语分测验和5个操作分测验的标准分数相加，便可得到全量表总分。之后，参照被试所属的年龄组常模，将上述三个量表分别转换为平均数为100、标准差为15的离差智商分数，就得到了言语智商、操作智商和总体智商。由此不但可以评价一个人的一般智力水平，也可以了解他在不同能力方面的差异。

### 3. 瑞文标准推理测验[Raven's standard progressive matrices test]

瑞文测验属于非文字智力测试，有三个版本，一个是1938年出版的标准推理测验，另外两个于1947年编制，分别是彩色推理测验和高级推理测验。彩色推理测验适用于5~11岁儿童和智力落后成人，高级推理测验则用于高智力水平的成人。测验具有较好的信度和效度。

瑞文测试主要测量人的抽象思维能力或者说推理能力。其特点是：(1) 全部使用图形作为测验的内容，可以进行跨文化的研究。(2) 主要是测试逻辑推理能力，而不是语言能力(被测试者进行一系列的数字推理)(3) 得到的分数只是一个标准分，不是一个智力的分数。

#### 13.3.3.3 儿童智力标准化测评的实施

如今，标准化的智力测验诞生于法国，弘扬于美国并且已应用于全世界各主要国家，智力测验的内容已从简单的认知测验发展到复杂的智力测验，特别是对思维的测验。当前，在新的智力结构理论的影响下，智力测验开始着重于对元认知成分的测验，并且常与智力训练结合起来进行，或由个人测验发展到团体测验。智力测验也广泛地应用于工业、教育、医学、军事、文艺和人才选拔等方面。但是，由于智力活动的复杂性，智力测验还有待于不断改进和完善。

对儿童智力的标准化测评是一项严肃的工作。尽管人们对其测验方法仍存在各种争议，但是，为了解儿童的智力状况，为了帮助儿童更好的发展，智力测验依然是当今探知一个人智力素质的一

个有用的工具，应该有条件的采用并认真对待，以充分利用其有效的一面，而避免其负面影响。这里，仅提出几点希望。

### 1. 实施测评时要注意的一些事项

① 采用量表进行测试的人员，一定要阅读其手册，按要求认真操作；② 主试者必须受过进行个体和团体测验的训练，掌握量表的正确测量技术—特别是提问技术，如，鼓励回答的技巧、书写回答的格式、记分的方法、记分的标准、原始分(粗分)换算标准分(量表分)的方法等，要能正确计算智商并能对结果进行科学的解释等；③ 测验材料要准备好，以方便测验时取用，使用能得心应手，不致紊乱，不影响进行时间；主试者要注意让被试者操作自然，避免对被试的操作带来不良影响；④ 测验时间要选择恰当，这是与被试者建立良好协调关系所必需的；被试者应在精力充沛、身体舒适、没有急事的时候来接受测验；⑤ 主试者应努力取得被试者的合作，尽量使他们保持对测验的兴趣。

### 2. 对测验要有清醒的认识

人类个体的智力发展确实存在差异：一是智力水平会随着个体年龄的增长而发生不同的变化。一般来说，智力的发展可以分为三个阶段：增长阶段、稳定阶段和衰退阶段。二是在智力发展的水平上，个体之间智力确实有高有低。在智力的结构上，个体之间确实存在不同差异。在智力成熟的时间上，个体之间确实有早有晚。

智力测验确实可以识别出不同智力发展状态的儿童。而智力测验的初衷也是希望通过智力测试找出有特殊需要的儿童，认为在正态曲线分布中，偏离正常水平的两个极端的群组均需要特殊的教育。特别是在学习过程应尽早发现处于低水平的儿童，以确保在他们早期对他们采取适合的教育。事实上，对于许多特殊的儿童，如何认识其智力状态，进而因材施教，才是智力测评最根本的目的。

但是，我们也要清醒的认识到，现有的智商测量法并不具备全面测量一个人—特别是儿童的智力的功能，也不能全面检测一个人智力发展的进程。而所谓“智商”，也不具备准确验证智力的效能。因为 IQ 分数高低经常要受到测验样本和常模的左右，受到被测者出身背景和教育“模式”的制约，也受到被测者练习和经验的影响。基于传统智力理论的机械的和片面的测试使它难以反映一个人的现有智力，更不用说用它去预测一个人未来的智力和能力的发展问题了。现有智力测验并不是公平、客观的。因为现有智力测验测量的主要是认知能力，只是人类智力或智能的一个方面。因此，所有的智力测验的结果只能作为我们考察儿童智力状况时的一个参考，若用它来作为评定一个人的标准或作为“贴标签”的依据则是错误的。

智力测验只是一种手段，一种工具。为了更有效的利用它，一方面，对现有的智力测试，我们应在理论和实施方法方面努力改进，尽量使其客观合理。智力测验无疑要受两个方面的支配：理论方面，依据不同的智力理论设计出来的智力测试量表通常只可用来测量智力的特定方面的性质；文化方面，智力测试量表的应用需要根据不同地区的文化作出相应的修改。一个标准的测试只对其相应的人群有效。因此，这种测试本身就需要被定期测评，因为测试的标准是会随着时代和社会环境的不同而变化的。另一方面，为了使评价更有效，使用其它一些必要的手段来证实智力测验的结果也是必要的。采用一些更讲究实际的评价方法，包括父母、老师等长期直接接触孩子的人的观察和看法，从孩子每天的行为来了解孩子的潜力和不足，也是一种更实用的评价方法。

即使真的需要解释测验结果以及要应用测验结果去帮助孩子解决学习困难时，也应注意：① 要谨慎地利用智力测验的分数作出解释，且不要纯粹地为了满足管理目的而去解释。② 不要把智力测

验认为是绝对可靠的工具，最终的测验应该在于学生的真实表现情况。③ 教师不仅应该认识到智力分数的表观价值，而且应该试图通过智力分数去对学生未来的潜质做一个判断。

教师，作为人类灵魂的工程师，被授予了运用各种各样的教育资源来最大化学生潜能的责任和权利。这些职责只有在教师对资源的运用有了整体理解后才能正确地行使。教师需要对学生的个体需要很敏感，并且应该准确了解学生的个体差异。教师最大的挑战之一就是处理同一班级中学生的个体差异。只有在针对不同学生的特点采用不同的引导方法时，教育对学生才是真正有用和有效的。教育资源的稀缺也许不允许我们根据不同学生的不同需要来培养学生，但尽量做到这一点应是教育发展的方向。

## 13.4 成人职业能力和素质测评

### 13.4.1 职业能力与素质

什么是能力？能力是一种直接影响人的活动效率，使活动顺利完成的个性心理特征，是顺利实现某种活动的心理条件。能力有一般能力，即在各种活动中均可表现出来的能力；也有特殊能力，即在某种专门活动中表现出来的能力。其中，思维能力是一般能力的核心；认知能力是个体接受信息、加工信息、运用信息的能力；操作能力是操纵、制作和运动的能力；模仿能力是通过观察别人的行为、活动来学习各种知识，然后以相同方式予以表现的能力；创造能力是可产生新思想，发现和创造出新事物的能力；社交能力是人们在社交活动中表现出来的能力。能力既包括实际表现出来的能力（ability），也包括个体潜在的能力（capacity）。

能力存在个体差异。能力的个体差异可表现在能力的类型、能力发展的水平和能力表现的早晚等多个方面。对于成人来说，人们关心的主要是其能力类型以及其可胜任特定工作的能力。这些能力可表现在一般能力的差异，如在知觉能力、记忆能力、言语能力和思维能力等方面，也表现在特殊能力的差异方面。比如，在知觉方面，有的人属综合型—知觉具有概括性和整体性，但分析能力较弱；有的人属分析型—知觉的分析能力较强，对事物的细节能清晰地感知，但对事物的整体知觉较弱；有的人属分析综合型—知觉兼有上述两种类型的特点。在记忆方面，根据个人记忆材料的方法的不同可以分为：视觉型、听觉型、运动型、混合型；根据个人识记不同材料的效果和方法的不同可以分为：直观形象的记忆型、词的抽象记忆型、中间记忆型。在言语和思维方面，有的人属生动的言语思维型—在言语和思维中有丰富的形象和情绪因素；有的人属逻辑联系的思维言语型—思维和言语是概括的、逻辑联系占优势；有的人属中间型。特殊能力的差异是指完成特定活动时必须具备的能力的差异，这对从事不同工作十分重要。例如，达芬奇在十几岁时到一个寺院里游玩，看到许多壁画和雕塑，回家后他全部默画出来，不仅轮廓、比例、细节一样，而且色彩明暗也很逼真；法国风景画家罗兰不是对实物风景作画，而是回家后根据视觉表象来画风景画；贝多芬在完全耳聋后，仍能根据听觉表象创作出第九交响曲；这些都是特殊的才能。特殊的才能也可以由能力的不同组合来保证。例如，同是音乐成绩优异的儿童，一个可能具有强烈的曲调感和很高的听觉表象能力，但节奏感弱；另一个可能具有很好的听觉表象能力和强烈的节奏感，但曲调感较弱；第三个可能具有强烈的曲调感和音乐节奏感，但听觉表象能力较弱。他们三人都有音乐才能，但音乐才能的构成成分存在着差异。

职业能力是从事具体职业所要求具有的能力。既包括一般能力，也包括特殊能力。在更多的时候，我们用职业素质来表达一个人从事某一特定职业所应具有的各种基本能力、心理素质和道德品

质等。不同的职业，要求的职业能力和素质也各不相同。我国教育部考试中心将职业分为：研究型、技术型、经营型、社会型、自然型、事务型、艺术型。Roe根据自己的理论，将各种职业分成八类，如图所示。



图13.4.1 Roe提出的职业分类表

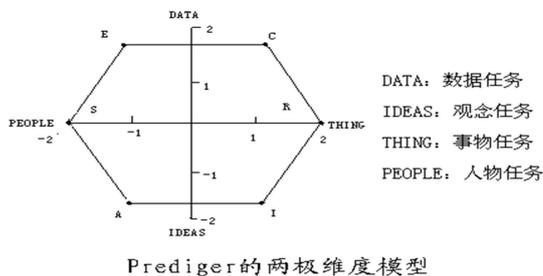


图13.4.2 Prediger关于职业类型的两极维度模型

**Prediger则提出关于职业类型的两极维度模型。**Prediger 认为，职业的基本任务可概括为分别处于两个维度的四种典型类型。其中，“数据任务”（DATA）主要通过事物、记录、文件、数字及系统程序打交道，从而服务于人们的日常消费和服务消费，如图书管理员、交通管理员的主要工作符合这一特点；“观念任务”（IDEAS）主要涉及概括、理论、知识、洞察以及以新方法呈现事物，科学家、作曲家、哲学工作者的主要工作符合这一特点；“事物任务”（THING）主要涉及仪器、材料、工具、生物机制等，建筑工人、实验室技术人员、司机的主要工作符合这一特点；“人物任务”（PEOPLE）主要与人打交道，涉及看护、游说、娱乐、训练等，教师、社会工作者、职业咨询师的主要工作符合这一特点。

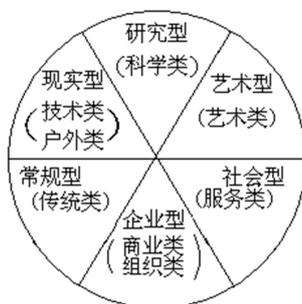


图13.4.3 Holland 职业与个性分类表

不同类型的人可能适合不同类型的工作，而不同类型的工作也需要不同类型的人。例如，霍兰德（Holland）将个性划分为六种基本类型，并列举出其特点，以及与之相匹配的职业类型。它们分别是：（1）现实型（Realist）的人，偏好与物体打交道，喜欢摆弄和操作工具、机械、电子设备等具体有形的实物，不喜欢和人打交道；他们喜欢选择机械、电机、制造等领域的职业。（2）研究型（Investigative）的人，喜欢用头脑依自己的方法来解决问题并追根问底，他们喜欢提出新的想法和策略，但对实际解决问题的细节较无兴趣；他们多喜欢从事数理、生化等领域的研究工作。（3）艺术型（Artistic）的人，喜欢用文字、音乐、色彩等不同的形式来表达情绪或美的感受；喜欢创造，不喜欢受束缚；他们喜欢从事音乐、写作、戏剧、绘画、设计等领域的工作。（4）社会型（Social）的人，关心自己和别人的感受，喜欢倾听和了解别人，愿意付出时间和精力去解决别人的冲突，并帮助他人的成长；他们喜欢从事教育、咨询、慈善福利等领域的工作。（5）企业型（Enterprising）

的人，多希望拥有权力去改善不合理的事情，善用说服力和组织能力，希望自己的表现被他人肯定；喜欢从事管理、行政等领域的工作。(6) 事务型 (Conventional) 的人，做事规矩而精确，喜欢按部就班、精打细算；不喜欢改变或创新，也不喜欢冒险或领导；他们多喜欢从事文书事务、金融、统计等领域的工作。

### 13.4.2 职业能力和素质的测评

什么是职业能力和素质测评？职业能力和素质测评是测评者运用一些科学的手段与方法在较短时间内快捷了解与评价某一个体的职业能力和素质的活动。例如，通过让被测评者回答一些问题，了解其知识、技能水平的测试或面试，通过让被测评者参与某种讨论，了解其沟通与思维的能力等。

为什么要进行职业能力和素质测评？其必要性包括：① 为职业选择人才提供帮助。通过合适的测验和评估，可以预测人们从事各种活动的适宜性，从而可以提高人才选拔和职业训练的效率。② 为人才选择职业提供帮助。通过合适的测验和评估，可以了解人的特长和特定素质，或对其因材施教，或为其推荐和安排合适的职业，以做到人尽其才。③ 素质和能力缺陷辅助诊断。可以通过测试辅助诊断各种智能缺陷和精神疾病等，如性格弱点和心理障碍等。④ 素质和能力状况辅助评价。可以通过测试评估人们在能力上的差异、人格的特点以及相对的长处和弱点，以利于有目的的培养和锻炼自己。

能力和素质测评就是通过一系列测试和考察来确定一个人能力和素质的广度和发展水平。我国自古就有采用一定的手段和工具来评定人的能力和素质等的记载。例如，尧在暴雨大作之日置舜于山林大泽之中，以测其辨别方向的能力，这可以说是最早的自然迷津测验。西汉的杨雄曾用言语和书法的速度来判断人的智慧。此外，还有人用猜谜语、作对联、解九连环、拼七巧板等方法测量人的能力。但是这些方法都比较简单，也不够严谨和系统。为了满足社会的需要，20 世纪初，人们开始制定出各种量表来测量人的能力和素质，其目的在于以标准化和数量化的方式将人的能力和素质精确地表示出来。能力和素质测验可以有不同的类型，如：智力测验、特殊能力测验、创造力测验、心理素质测验等。如今，一些专业化的测评系统可测项目见图13.4.4和表13.4.1所示。

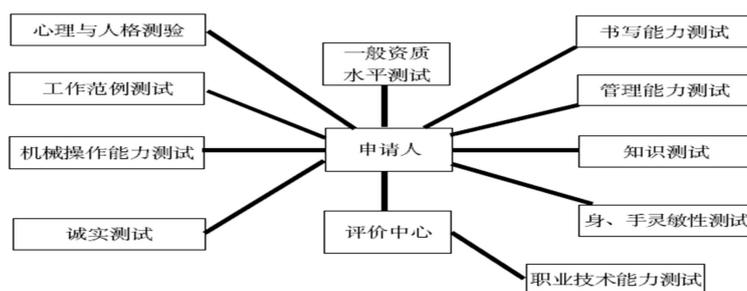


图13.4.4 测评系统可测评项目

表13.4.1 职业能力与素质测评系统可选测评项目表

序号	测验类别	测验名称	测验英文名	题目数量	答题时间
1	基本潜能	机械推理能力测验	Mechanical Reasoning Ability Test	30	15
		逻辑推理能力测验	Logical Reasoning Ability Test	60	25
		知觉速度测验	Perceptual Speed Test	96	25
		空间关系能力测验	Spatial Relation Ability Test	36	15
		语言能力测验	Verbal Ability Test	60	25
		资料分析能力测验	Data Analyst Ability Test	20	20
		数字运算能力测验	Numerical Operation Ability Test	40	15
2	招聘筛选	普通员工招聘筛选测验	Employee Screening Test	96	30
3	智力测验	瑞文智力测验	Raven Standard Progressive Matrices	60	30
		华瑞智力测验	HR Intelligence Test	65	25

序号	测验类别	测验名称	测验英文名	题目数量	答题时间
4	个性特质	气质测验	Temperament Test	60	20
		爱德华个人偏好测验(EPPS)	Edward's Personal Preference Schedule	225	45
		卡特尔16种人格因素测验(16PF)	Cattell's 16 Personality Factor Questionnaire	187	30
		加利福尼亚个性调查表(CPI)	Carifornia Psychology Inventory	480	60
5	行为风格	领导权变模式测验	Convergent Leadership Model Test	58	10
		领导行为模式测验	Leadership Behavior Model Test	35	10
6	动机态度	工作满意度测验	Job Satisfaction Test	56	20
		个人需要测验	Personal Needs Test	50	20

序号	测验类别	测验名称	测验英文名	题目数量	答题时间
7	职业发展	职业价值观自我测验问卷	Career Value Questionnaire	52	15
		一般职业能力倾向自评测验	Career Aptitude Self-assessment	45	20
		职业兴趣自我评定量表	Career Interest Self-assessment	72	25
		职业倦怠量表	Chinese Maslach Burnout Inventory	15	8
8	职业能力评估	销售潜能测验	Sale Potential Test	155	30
		管理能力测验	Management Ability Test	72	40
		IT能力倾向测验	IT Ability Test	40	25
		HRM从业者胜任力自评问卷	HRM Competency Self-assessment	55	20
9	组织诊断	组织承诺问卷	Organizational Commitment Questionnaire	15	10
		工作气氛量表	Organizational Climate Questionnaire	30	8
		组织公民行为量表	Chinese Organizational Citizenship Scale	20	10
		工作价值调查表	Work Values Inventory	25	8
10	满意度	员工满意度测验	Employee Satisfaction Test	39	10

### 13.4.3 职业素质与能力的测评的内容

目前,一些职业素质与能力测评的内容主要包括:

#### 1. 相关知识测评

一定的知识水平是从事一定职业所要求的基本条件之一,也是考查一个人的专业能力的一个基础。对知识的测评包括基础知识和专业基础。基础知识测评主要考察被测者从事某一职业时应知应会的最基本知识,如基本常识等,它可涵盖政治、经济、法律、管理、人文、科技等多个方面,重

点考察的是被测者在这些方面知识的广度；专业知识测评主要考察被测者从事某一职业时所应掌握的专业知识的深度。

## 2. 相关能力和能力倾向测评

能力是从事一定职业的核心要求。对能力的要求包括对现有能力的要求和对一定能力倾向[性向、潜在能力]的要求。能力有一般能力和特殊能力，因而对能力和能力倾向的测试也包括一般能力[能力倾向]测评和特殊能力[能力倾向]测评。测评的一般能力可包括：一般学习能力、语言能力、数字能力、空间能力、图形知觉能力、文书知觉能力、动作协调能力、手指灵活性、操作灵活性等。其中，智力测评是最典型的一般能力测评。智力是人类学习和适应环境的基本能力，包括观察能力、记忆能力、想象力和思维能力等，智力测验的项目也就包括了数字、推理、判断、分析、理解能力等。智力的高低常用智商来表示（IQ）。一般认为，IQ在130以上为超常水平，IQ高适合担任难度高的技术工作；IQ在70—130之间为正常水平。有些研究认为，职业人员的IQ并非越高越好；在一个团队中，所有人员的IQ都高并不一定有利于工作。标准化的智力测验包括韦氏测验、瑞文测验、斯坦福-比奈智力测验等。测验可测量人的多方面的能力。我们主张一般能力测评应注重考查一个人的思维能力。其测评的标志包括：回答问题层次是否清楚；论述问题是否周密；阐明论点论据是否连贯；运用知识分析判断问题的能力是否灵活准确；言语理解与表达是否清晰等。

特殊能力是指某些人所具有的他人所不具备的能力。特殊的职业需要特殊的能力。职业能力的测评依不同的职业要求有很大的不同。其测评项目可包括对视力、音乐、艺术、飞行等方面能力的测评；对机械、空间、运动、体能等方面的机能测评等。一些典型的特种能力测验包括霍兰德职业倾向测试、威廉斯创造能力测验、明尼苏达操作速度测验（Minnesota Rate of Manipulation Test）等。

## 3. 创新能力的测评

对创新人才的需求也促进了创新能力的测评。由于创新能力是一种特殊的综合能力，因而评价一个人的创新能力是一件很困难的事。目前，人们对创新能力的测评主要是对创造性思维的测评。创造性思维可表现在多个方面，比如，在常规性工作方向上是否具有不同于别人的新思路（创新思路），在工作方法上能否采取新的措施（创新方法），对已完成工作是否具有新的见解（创新见解）等。对创造性思维的测评也有多种方法。目前常用的方法有3类：（1）以发散思维为指标的创造力测验。这是运用得最多的一种测验形式。这类测验大多以发散思维为指标，从流畅性、变通性、独特性等几方面评分。著名的测验有：建立于吉尔福特三维智力模型理论基础上的南加利福尼亚大学发散思维测验；托兰斯创造性思维测验；盖策尔斯和杰克森门编制的芝加哥大学创造力测验等。（2）以创造性个性为指标的创造力测验。这类测验以创造性个性为指标，考察被试是否具有好奇心、独立性、恒心、适应性、自信心等创造性个性特点。（3）以创造性成果为指标的创造力测验。这类测验通过创造性成果来评定个体的创造能力。不同领域的创造性成果是不同的，如科学论文、新发明、音乐作品、诗歌等都是不同领域的创造性成果。从成果入手来评价创造力相对来说较为直接而且公正。

许多心理学家研究了创造性和实际创作作品之间的关系。瓦拉齐（M. A. Wallach）等人曾以500名大学生作为被试，发现思维的流畅性和创造作品之间有明显相关。思维流畅性能够预测许多领域中的成就。创造力测验是心理测验适应时代需要的一个新动向。不过，当前的几个创造力测验的信度比一般的智力测验还低。它们仅能在一定程度上预测一个人的创造成就的大小。

#### 4. 职业心理素质测评

职业心理素质测评属心理测评，主要是针对与特定的职业要求相适应的个性心理素质进行测评。个性是个体所具有的独特而稳定的心理特征，这些特征决定了个体在适应环境时的行为模式。个性包括气质、性格、兴趣、爱好、价值观等。心理学家霍兰德曾将个性划分为六种基本类型，并列举出其特点，以及与之相匹配的职业类型。

兴趣是在一定需要的基础上，在社会实践过程中形成并发展起来的一种对事物或活动喜欢与不喜欢的相对持久的心理倾向。职业兴趣是兴趣在职业选择活动方面的一种表现形式，是指对职业或具有职业特征的活动的选择性态度和积极的情绪反应倾向。

职业兴趣一般分为[Kuder]：

户外 (Outdoor)：喜欢大部分时间能让你待在户外的活动，这些活动通常与动植物打交道。

机械 (Mechanical)：喜欢与机器和各种工具打交道的工作

计算 (Computational)：喜欢与数字打交道的工作

科学 (Scientific)：喜欢发现新事实、解决问题

说服 (Persuasive)：喜欢与人们打交道，参加各种集会，兜售自己的观点和想法，推动方案的执行的工作

艺术 (Artistic)：喜欢有创造性的工作，包括在设计、颜色、形式和材料上进行各种有吸引力的创新

文学 (Literary)：喜欢阅读和写作

音乐 (Musical)：喜欢参加音乐会，演奏乐器、唱歌，看各种关于音乐和音乐家的东西

社会服务 (Social service)：喜欢帮助他人

公务 (Clerical)：喜欢要求精确度和准确性的任务

职业心理素质测评包括多个方面。其中，职业兴趣测验主要测试一个人对某一职业是否有兴趣。比如，有一个考查经营取向和社交取向两种类型的职业兴趣的测试如下：

本测试旨在帮助你选择自己更喜欢的或者更适合自己的职业，使你找到满意的工作。本测验的每一个题目都给出一种活动、或一种技能、或一种职业，请你依自身情况选择。

你是否乐于从事与人打交道的工作？请你依自身情况选择：

1、策划企业活动；2、参加联谊活动；3、领导促销活动；4、调节邻里纠纷；5、销售经理；6、为公益事业做义务宣传；7、企业经营顾问；8、当教师；9、投资商；10、做职业咨询顾问；11、市场行情策划；12、安慰别人。

答案：A 非常不喜欢；B 稍有不喜欢；C 无所谓；D 稍有喜欢；E 非常喜欢。

说明：经营取向(1, 3, 5, 7, 9, 11)：6-12分为较低，13-24分为中等，25-30分为较高；

社交取向(2, 4, 6, 8, 10, 12)：6-12分为较低，13-22分为中等，23-30分为较高。

分数越高，表明你越喜欢该类型的工作。以上问题测试了经营取向和社交取向两种类型的职业兴趣。

分析：经营取向的个体喜欢诸如推销、服务、管理类型的工作。这种取向类型的人往往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。这种人喜欢同人和观念而不是与事务打交道。他们热爱交际、冒险、精力充沛、乐观、和蔼、细心、抱负心强。社交取向的个体喜欢社会交往性工作，如教师、咨询顾问、护士等。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。

气质测评也是常见的一类职业素质测评。气质是个体中与神经过程的特性相联系的行为特征，是个体心理活动和外显动作中表现的某些关于强度、灵活性、稳定性与敏捷性等方面的心理特征的综合。巴甫洛夫认为，人的神经活动有兴奋与抑制两种过程，分为三种特征：活动强度、均衡性与灵活性。希波克拉特则将气质分为如下类型：

气质类型	特 性
胆汁质	直率 热情 精力旺盛 情绪易于冲动 心境变化剧烈 具有外倾性
多血质	活泼 好动 敏感 反映迅速 喜欢与人交往 注意力易发生转移 兴趣容易变换 具有外倾性
黏液质	安静 稳重 反应缓慢 沉默寡言 情绪不易外露 注意力稳定难于转移 善于忍耐 具有内倾性
抑郁质	孤僻 行动迟缓 善于观察细小事物 情感发生较慢但持续很久 体验深刻 具有内倾性

图13.4.5 希波克拉特的气质类型特征

不同年龄阶段的人的气质存在着很大的差异；个体大部分为两种或更多种气质的混合类型；对气质进行测评的例子有斯特里劳气质调查表等。

对价值观的测评（有的也包括品德）属态度测评。对价值观进行测评的例子有斯普兰格关于价值的类型评定等。

类型	表 现
理论型	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
经济型	强调事物的实用性 凡事以有效和实惠为尺度
审美型	总是从形式和和谐中寻找最高价值 以文雅、优美、对称和恰当的标准去判断每一种经验
社会型	利他与仁慈 在实际生活中表现出宽容 富有同情心和无私等品德
政治型	热衷于个人权利、影响力和声望 在政治以外的领域中也希望通过竞争、奋斗而获取名誉和地位
宗教型	理想理想主义者 工作与生活中以自己的信仰与理想为准则

图13.4.6 斯普兰的格价值类型分类

### 13.4.3 职业素质与能力测试的主要方法

职业素质与能力测试大多属心理测试。心理测试是运用一系列心理学方法来测量被试的能力水平和个性方面差异的一种科学方法。其基本假设是：① 个体的行为表现，都是个体的心理素质在特定环境下的特定表征。② 个人的素质是相对稳定（不是绝对）的，各个个体不尽相同，它可以综合不同环境下的刺激，使个体在不同的刺激下作出一致的反应。

心理测试的方法包括纸笔测试[笔试]、投射法和心理实验法等。所谓投射法就是让被试者通过一定的媒介，建立起自己的想象世界，在无拘束的情景中，显露出其个性特征的一种个性测试方法。心理实验法则是指有目的严格控制，或者创造一定条件来引起个体某种心理活动的产生，以进行测量的一种科学方法。

心理测验是收集有关个别差异的资料的一个简便易行而又较为可靠的方法。心理测验的注意事项包括：（1）坚持守密原则。私人测验材料一般要保密，私人测验结果一般也要保密；知情同意（Informed Consent）：未经许可不得向任何人或机构公布带有受测者姓名标识的测验结果，除非有预先约定。（2）客观看待测验结果。不可绝对化，不可过分怀疑，不可过分依赖；（3）避免对个体贴标签。不可草率下结论，提倡追踪测验与横向比较，必要时采用最轻烙印标签（least stigmatizing label）。

测评时要注意：确立测评指标体系科学合理；职位分析得出的测评要素要准确恰当；根据测评要素建立的测评标志与标度要符合实际要求；根据指标体系确立的具体项目，根据具体项目确定的赋分标准要能反映真实情况。

职业素质与能力的测试，目前常用的方法是笔试、面试、情景模拟和标准测试等。

### 1. 笔试

笔试是以（广义）纸笔形式对被试的能力、性格及行为特征进行考核和评价的一种方法。是提出问题让被试作出书面回答。其题型包括主观题，也包括客观题。其特点是：公平、费用低、迅速、简便；测试多注重思维的结果，而不是思维的过程；测试易标准化，且易于评定。

笔试最适合对一个人所掌握的知识进行考核。知识考试是通过笔试的形式对被试的知识广度、知识深度和知识结构进行了解的一种有效方法。知识考试可做百科知识式的考试，其考核内容广泛，以了解被试知识的广度；也可做专业性的知识考试，以了解被试专业知识掌握的水平。

知识考试的操作要点在于：一是试卷的设计，要求目标明确、内容全面、强调运用；应尽量建好题库，请专家出题，严格操作；二是考试的安排，要求公平公正、可使被试发挥出应有水平；三是阅卷要求标准尽量客观统一。

### 2. 面试[面谈与观察]

面试是主试通过观察或与被试的交流来对被试的能力、性格及行为特征进行考核和评价的一种方法。是一种要求被试用口头语言来回答主试提问，从中了解被试心理素质和潜在能力的测试方法。面试可包括平时面谈（聊天）、正式面试、随机问答和论文答辩等。其特点是：重过程；不易标准化。

面试多用于单位招聘。面试的目的，对求职者，是在限定时间内向招聘人员推销自己，同时也通过面试了解所应聘公司及工作性质，看看是否与自己的期望相符。对招聘人员，有查核：通过问答，了解应聘者提供的资料及推荐书上的意见是否可信；有观察：面对面观察申请人的仪容、态度、谈吐等；有测验：用各种办法评估应聘者的性格、才能及知识等；有判断：考虑申请人是否有能力及诚意担当此职，并是否适合在该公司工作。

结构化面试是面试中常见的一类。所谓结构化面试是指命题、实施结果评定等环节均按事先制定的标准化程序进行的面试，因而亦称标准化面试。结构化面试中的标准化主要体现在下面几个方面：（1）以工作分析为基础。（2）每一次面试之前根据实际情况选择该职务测评要素中最重要的、可测量的部分命题，且题目均事先给出参考答案及评分标准。（3）对所有的应试者都提相同的一组问题。（4）考官须经过专门的培训。（5）有统一的评分标准和评定量表。（6）每个应试者的面试时间相同（一般30—40分钟）。结构化面试的命题要求符合测评要素的要求，并坚持思想性、科学性、理解性、开放性的命题原则。

面试中的提问方式有一定技巧。常见的提问方式包括：（1）连串式提问。向面试者提出一连串相关的问题，要求应试者逐个回答。这种提问方式主要是考察面试者的反应能力、思维的逻辑性和条理性。例如：“你在过去的工作中出现过什么重大失误？如果有，是什么？从这件事本身你吸取的教训是什么？如果今后再遇到此类情况，你会如何处理？”（2）开放式提问。所谓开放式提问，就是指提出的问题应试者不能使用简单的“是”或“不是”来回答，而必须另加解释才能回答清楚。因此，主考官提出的问题如果能引发面试者给予详细的说明，则符合开放式提问的要求。（3）压迫式提问。一般来说，主考官要尽力为应试者创造一个亲切、轻松、自然的环境，以使应试者能够消除紧张、充分发挥。但有些情况下，主考官会故意制造一种紧张的气氛，给应试者一定压力，通过

观察应试者在压力情况下的反应，来测定其反应能力、自制力、情绪稳定性等等。例如：“这次公务员考试，很多人都托了关系，听说你也走后门了。”（4）假设式提问。在这种提问中，主考官为应试者假设一种情况，通过应试者的反应和回答，来考察应试者的应变能力、解决问题的能力、思维能力。如：“如果你是那个肇事的司机，你会怎样处理？”（5）清单式提问。除了提出问题外，还给出几种不同的可供的选择的答案。目的是鼓励应试者从多种角度来看这个问题，并提出了思考问题的参考角度；比如，“你所在企业中最主要的问题是什么？营业额、管理、产品质量差还是其他？”（6）封闭式提问。一种可以得到具体答案的问题。这类问题比较简单、常规，涉及范围较小。关于下面的一些情况常用封闭式提问：工作经历（包括过去的工作职位、成就、工作成绩、个人收入、工作满意与否以及调动原因）、学历（包括专业、学习成绩、突出的学科、最讨厌的学科、课程设置等）、早期家庭状况（包括父母的职业、家庭收入、家庭成员等）、个性与追求（包括性格、爱好、愿望、需求、情绪、目标设置与人生态度等）等。

面试应注意的问题有：（1）既要注意制造和谐的面试气氛，同时也要注意尽快转入正题的技巧；（2）尽量避免面试人员主观因素干扰。如，首因效应；定势思维；对比误差；雇用压力；注意非语言行为身体魅力；晕轮效应；个人喜好等。

### 3. 情景模拟

情景模拟测试是把被测试者置于一个模拟的工作情景中，从而观察和评价他们在模拟的工作情景的压力下的行为。它最初由美国战略情报局在二战期间发明并使用，他们的任务是派遣特工人员去敌后进行活动，他们的挑选程序旨在设计出一个具有高度压力的情景，让特工候选人在此情景中工作，观察他们的行为，以此来预测他们到敌后去活动的的能力。

现在的情景模拟主要根据被试可能担任的职务编制一套与该职务实际情况相类似的题目，将被试安排在模拟的、逼真的工作环境中，要求被试处理可能出现的各种问题，用多种方法来测评其心理素质和潜在能力等。其特点是：①设计复杂，准备时间长，但正确度较高；②测试费用较高；③适用于对高级管理人员与特殊人才的招聘。

情景模拟测试的内容与形式可包括：（1）公文处理。公文处理是给被试不同的处理文件，让被试处理，从而反映被试不同的处理能力。如：办事效率；利用工作条件的分析能力；发现问题的能力。公文处理测试对被试的要求是：参加公文处理的被试应该像准备担任该职位的正式工作人员一样从事各项工作，仔细审阅各种公文的内容，解决各种问题，回答各种问题，组织，安排进度和计划等。其设计时要注意：有确定的评价标准；公文内容与测试目的要相符合；时间要适当；环境要逼真；指导语要准确。（2）与人谈话。包括：①与人电话谈话—可用于测试被试心理素质、文秘修养、口头表达能力、处理问题能力。②负责接待—可用于考察被试待人接物的态度、快速处理问题的能力。③拜访有关人士—可用于考察被试公关的技巧、语言表达能力，热情的程度，应付各种困难的能力。其设计时要注意：要有明确的测试的目的和内容；要选择合适的谈话方式；评价标准要一致；为被试提供必需的资料。（3）无领导小组讨论。是测试被试领导和组织技能的一种非常有效的方法。可通过观察评价被试的领导欲望、主动性、说服能力、口头表达能力、自信程度、抵抗压力的能力、人际交往能力等。设计时要注意：恰当确定小组成员；恰当确定讨论时间；安排好讨论议题；准备好必要的资料；有确定的评价标准。（4）角色扮演。要求被试扮演特定的管理角色来处理日常的工作事务，以此来观察被试的多种表现，了解其心理素质和潜在能力。设计时要注意：有确定的正确的评价要素；有规范统一的指导语；有设计好的活动细节。（5）即席发言。可用于测

试被试的快速思维能力、理解能力、思维的发散性、语言表达能力、言谈举止、风度气质等心理素质。设计时要注意：确定好题目；制定好评分标准；评价时抓住主要的心理素质。

#### 4. 标准测试

标准测试通常是在一定的理论指导下采用科学的方法制定的测试。如智力测试的韦氏测验、瑞文测验、比奈-西蒙测验、斯坦福-比奈智力测验；特殊能力测试的霍兰德职业倾向、威廉斯创造能力测验；升学考试的标准化测试等。标准测试编制理论基础较强，内容效度和效标效度较高，测试结果具有较高的可信度。

一些可以进行标准化测试的心理测试项目和标准化测试方法见图13.4.7。

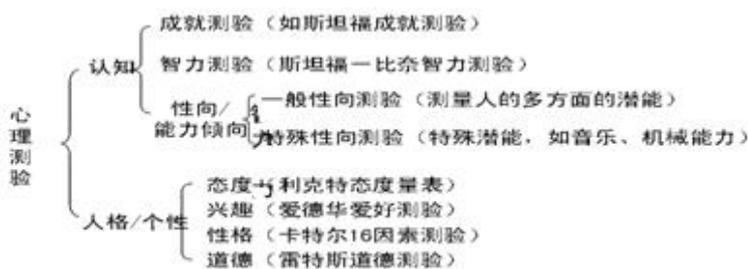


图13.4.7 一些心理测试项目和标准化测试方法

标准测试可用于一般能力测试和特殊素质测试。一般能力测试现在主要使用斯坦福-比奈智力量表和韦克斯勒智力量表。特殊能力和素质测试的标准测试则相对较弱。由于特殊能力和素质测试要测试被试所具有的他人所不具备的某种素质或能力。内容较杂，标准化并不容易。比如，创造力的测试，目前已有：

(1) 南加利福尼亚大学发散性思维测验。是美国南加利福尼亚大学的吉尔福特和他的同事编制的一套用于测量发散性思维能力的测验模式。测验的项目有：语词流畅性、观念流畅性、联想流畅性、表达流畅性、非常用途、解释比喻、用途测验、故事命题、事件后果的估计、职业象征、组成对象、绘画、火柴问题、装饰等。前10项要求言语反应，后4项则用图形内容反应。测验适用于中学水平以上的人，主要从流畅性、变通性和独特性记分。例如，“火柴问题”是要求被试移动规定数目的火柴，组成特定数目的正方形和三角形。

(2) 托兰斯创造思维测验。是美国明尼苏达大学的托兰斯 (E. D. Torrance) 等人编制了另一个著名的创造力测验，该测验分为3套，包括语词创造思维测验、图画创造思维测验和声音语词创造思维测验，共有12个分测验。为了减少被试的心理压力，他们用“活动”一词来代替“测验”。测验时根据4个标准评分：流利（中肯反应的数目）、灵活（由一种意义转到另一种意义的数目）、独特性（反应的罕见性）、精密（反应的详细和特殊性）。被试从整个测验中得到一个总的创造力指数，代表个体的创造性思维水平。该测验适用于从幼儿到研究生的文化水平，普遍采用集体测试的方法，对于小学4年级以下学生一般用个别口头测试。

(3) 芝加哥大学创造力测验。是美国芝加哥大学的盖泽尔斯 (J. W. Getzels) 和杰克逊 (P. W. Jackson) 等人根据吉尔福特的思想对青少年的创造力进行深入研究的基础上，在20世纪60年代编制的。这套测验包括5个项目：语词联想测验、用途测验、隐蔽图形测验、完成寓言测验、组成问题测验。这套测验适用于小学高年级到高中阶段的青少年，适用于团体测验，并有时间限制。

这些测试有一定作用，但其权威性尚不足。

再如，对个性的标准化测试，也有：

(1) 明尼苏达人格测验(MMPI)：有MMPI的分量表，用一组量表进行诊断和人格分析，有MMPI的分数的解释，可将测试结果转换成标准分，并绘制在人格剖析图上。MMPI可协助医生对病人的精神状况作出诊断并确定病情的轻重，对疗效判定和病情预后有一定参考价值。

(2) 卡特尔16项人格因素问卷(16PF)[卡特尔16种人格因素量表]，包括：

- |                            |                        |
|----------------------------|------------------------|
| A. 乐群 (Warmth)             | L. 警戒 (Vigilance)      |
| B. 聪慧 (Reasoning)          | M. 幻想 (Abstractedness) |
| C. 稳定 (Stability)          | N. 隐私 (Privateness)    |
| E. 恃强 (Dominance)          | O. 忧虑 (Apprehension)   |
| F. 兴奋 (Liveliness)         | Q1. 变通 (Change)        |
| G. 规则 (Rule-consciousness) | Q2. 独立 (Self-reliance) |
| H. 敢为 (Boldness)           | Q3. 求全 (Perfectionism) |
| I. 情感 (Sensitivity)        | Q4. 紧张 (Tension)       |

其常模取自美国全国上千人，据称其标准差2.4，内部一致性系数为0.64。

职业兴趣和个性测试也有：

(1) 爱德华个人偏好量表 (EPPS)。包括：

- |                      |                          |
|----------------------|--------------------------|
| 1、成就 (Achievement)   | 9、支配 (Dominance)         |
| 2、顺从 (Deference)     | 10、谦卑 (Abasement)        |
| 3、秩序 (Order)         | 11、抚助 (Nurturance)       |
| 4、表现 (Exhibition)    | 12、求变 (Change)           |
| 5、自主 (Autonomy)      | 13、坚毅 (Endurance)        |
| 6、亲和 (Affiliation)   | 14、异性爱 (Heterosexuality) |
| 7、省察 (Intracception) | 15、攻击 (Aggression)       |
| 8、求助 (succorance)    |                          |

(2) 五因素量表 (NEO-PI)。包括：

1、神经质 (Neuroticism, N) 量表：测量情绪稳定性，识别容易有心理烦恼、不现实想法的个体。高分：烦恼、紧张、情绪化、不安全、忧郁；低分：平静、放松、不情绪化、安全、自我陶醉。

2、外倾性 (Extraversion, E) 量表：测人际互动、活动水平、刺激的需求。高分：社交、活泼、健谈、乐群、乐观、重感情；低分：谨慎、冷静、冷淡、退让、不爱说话。

3、开放性 (Openness to Experience, O) 量表：测对经验的积极寻求和欣赏，喜欢接受并探索不熟悉的事物的程度。高分：好奇、兴趣广泛、有创造力、非传统、富于想象；低分：习俗化、讲实际、兴趣少、无艺术性。

4、宜人性 (Agreeableness, A) 量表：评鉴某人思想、感情和行为方面在同情至敌对这一连续体上的人际取向的性质。高分：心肠软、信任、助人、容忍、易轻信、直率；低分：愤世嫉俗、多疑、不合作、易怒、好报复。

5、认真性 (Conscientiousness, C) 量表：测个体的组织性、持久性和动力性的程度，把认真的人和懒散的人区分开来。高分：有条理、可靠、勤奋、自律、有毅力；低分：无目标、懒惰、不检

点、意志弱、享乐。

但它们同样只能做为参考。

我们认为,对素质与能力的测试和评估需要综合考虑多个方面,也需要多种测试方法的运用。现以性格评定为例。性格评定是指对一个人的性格进行描述和测量。常用的性格的评定方法有:观察法、实验法、作品分析法、测验法、个案法等。其中,观察法主要通过被试的外部行为表现,了解被试的性格特征,对被试的性格进行鉴别。作品分析法主要通过被试的作业、试卷、日记等作品进行分析,对被试的性格进行鉴别。测验法主要运用人格测验量表对被试的性格进行测量,从而对其性格进行鉴定。人格测验可分为自陈测验和投射测验两大类。明尼苏达多相人格测验、卡特尔16种人格因素测验、艾森克人格测验等属于自陈测验;罗夏墨迹测验、主题统觉测验等属于投射测验。各种性格鉴别的方法各有优缺点,要综合地运用各种方法,相互印证,从而对被试的性格做出较全面、客观的鉴定。

### 13.5 关于智力和能力测评的模糊综合评估方法研究

#### 13.5.1 对心理素质、智力与能力进行综合评估的方法

素质、智力和能力都是心理量。它们应该是一类相对的概念,并且与所在的环境、文化、传统等有密切的联系,是一个人所特有的与其所处的特定环境中的其他人相比较而言的一类心理参量。离开了特定的环境来谈论心理素质、智力和能力,是没有意义的。在大多数情况下,智能是对主体在学习和工作中处理各种问题的能力的表征。作为对一个人所具有的智能的一种衡量方法,智力和能力的测评所强调的,应是人在其特定的环境中所拥有的,面对特定问题发掘、利用、以及转化环境所能提供的各种信息和资源的能力。由于主体所具有的用于学习、工作和问题处理的各种基本心理素质、知识和实际操作能力各不相同[它取决于主体的先天的基质与后天的训练],个人所从事的工作的性质和要处理的问题并不相同,要从事不同工作和处理不同问题所需要的具体的基本素质、能力和知识也就各不相同。因此,要统一评价所有主体的智力和能力水平是很困难的,也是不必要的。我们只能针对不同的工作和学习环境对个体素质的特定要求,对主体的特定的素质、智力、知识和能力水平在一定范围内针对某一类特定目的做出一个大概的评估。并且,由于评估时所依据的“测试数据”与实际工作时的具体问题处理是不一样的,故评估出的“水平”不一定就是实际工作时的水平。对一个人心理素质、智力和能力的评估只能是一种“参考”测度。

从“智能是对主体信息处理能力的表征”的认知出发,对智能的评估要解决以下几个问题:(1)要针对哪类特定的信息处理问题进行评估?(2)要处理这类特定的信息需要哪些特定的素质、知识和能力?(3)这些特定素质、能力和知识各对主体的特定的信息处理能力产生什么影响?是否可以量化?(4)从那些具体事例操作[行为]或问题解答中可以测出主体的那些特定的素质、能力和知识?

#### 1. 评估指标体系的确定

评估指标(Indicator of evaluation)是根据一定的评估目标确定的、能反映评估对象某方面本质特征的具体评估条目。指标应是可具体化的、可测量的、可行为化和可操作的目标,即应是可标称目标的观测点、测量点,应是可以通过对评估对象的实际观察获得明确结论的。评估指标体系(system of indicators used evaluation)是由不同级别的评估指标按照评估对象本身的逻辑结构形成的有机整体。它是衡量评估对象发展水平或状态的量表系统。评估指标体系是系统化的、具有紧密联系的、可反映评估对象整体状况的一组指标或指标的集合。

在设计评价指标体系时,应对以下一些方面着重进行考虑,以期得到的考察结果能更好的符合评估的初衷。① **一致性**。指标体系应该是一个层次结构,所谓一致性是指,既要使评估指标与评估目标一致,又要使下一层次的指标与上一层次的指标一致。② **可测性**。评估指标系统中末级指标应能够以可操作化的语言加以界定,它所规定的内容可直接测量或清楚的加以描述的,以获得确定的结论。③ **可比性**。评估指标必须反映所有评估对象的共同属性,是所有评估对象中拥有的共同的东西。④ **独立性**。同一层次内的指标必须各自独立的,指标之间不能相互重叠和包含,既不能存在因果关系,也不能从一项指标导出另一项指标。⑤ **完备性**。指标体系必须是一个完整、协调的系统,能够全面地、毫无遗漏地反映评估目标或上一层次指标。⑥ **可行性**。评价指标的数量和评价标准的高低都要适中。有足够的信息、人力、物力和切实可行的量化方法可资利用。

对不同的评估目的,应有相对不同的指标体系。即使对于相同心理素质和能力的评估,指标体系也要适应当时社会文化、经济特征和技术条件的不同要求。为使评价指标具有一定的代表性,特别是让各个指标在整个评价体系之中能体现出相应的作用,应利用各种机会对指标系统进行较为深入的调研,与不同类别的相关人员进行较为深入的探讨。在深入调研的基础上,确定出适应不同测评目标的指标体系。

## 2. 常模定制与评估方法

按道理说,一个人所具有的特定的智力和能力应有相对水平和绝对水平,但心理测量常被认为是一类相比较而言的东西,被认为是与所处环境中其他的人相对而言的。作为相对比较,我们不但需要知道某个个体的智力和能力,还需要知道在这个具体个体所处的环境中,其他个体的情况,这样才可以确定这个个体应该处在什么样的相对位置上。然而,由于人口数量和状况的原因,要确切的定位某个特定个体在其所处环境中的具体位置是极为困难的。同时,这种所谓的确切定位对人清醒的认识自己的位置也是没有太大的实际价值的。扎德认为,“当系统的复杂性日趋增长时,我们做出系统特性的精确而有意义的描述能力将相应降低,直到达到某一阈值,一旦超过它,精确性和有意义性将变成两个互相排斥的特性”。

心理测评需要常模。简单的说,常模(Norm)就是参照物。根据现代抽样调查理论,按照某种规则从预先设定的群体中抽出部分个体,对这些个体进行测量。由这些测量结果所构成的集合就称为常模。有了常模这个参照物就可以对其他特定的个体的测量结果进行比较分析,获得该个体所拥有的智力在整体上处于什么样的位置。因此,常模的确定就成为了心理测评中最重要的关键内容之一。如何选取样本并计算出常模,会对心理测评带来极大的影响。韦克斯勒认为智力具有正态分布的特点,因此他提出采用“离差智商”的理论来衡量个体的标准分数,从而获得被测个体在其所属群体中相对的智力程度。正态分布由于其形状类似一口钟,因而也称为“钟形分布”或者“钟形曲线”。事实上,现实中有许多现象是符合或基本符合正态分布的。1994年,美国作家豪斯汀(Richard J. Herrnstein)和穆里(Charles Murray)曾发表了在美国引起极大争议的著作:《钟形曲线:美国生活中的智能和阶级结构》(The Bell curve: Intelligence and Class Structure in American Life)。在这部著作中,豪斯汀和穆里利用正态曲线描述并分析了存在于不同种族、不同国家和地区内部的智力分布。在某种程度上,可以将他们的结论看作是韦克斯勒“离差智商”的一种引申和广度的应用。如果说,韦克斯勒的智商是建立在不同年龄段人群的基础上的,《钟形曲线》的作者是智商建立在不同种族、不同国家和地区的基础上。

常模样本的选取可以考虑将抽样样本作为选择对象的重要组成部分,这些排行相对而言有较高

的权威性和客观性，但是也往往会有一定的局限性。尽管在大部分情况下，常模基本上是符合正态分布的，但也可能由于各种原因(如样本抽取时的偏差等)造成常模的分布曲线与正态分布曲线距离较远。为使抽取的样本更加符合要求，常常需要对样本进行正态化操作。所谓正态化操作是指根据统计学理论，通过对分布曲线的平移、拉伸和变形，使分布曲线符合正态分布曲线的要求。显然，这种正态化操作必须是在不改变原有曲线上观测点相对位置的前提下进行。正态化过程是一种比较简单的过程，其方法是：(1) 分别对不同的一级指标被抽取的人在这个一级指标上的得分按照从小到大的顺序排序。(2) 计算每个原始分以下的个体的数量占总的常模中的百分比例。这个比例称为该原始分的百分等级。(3) 通过标准正态分布表确定每个百分等级的正态分数。

上述这个过程只是实现了原始分的正态化要求。为使不同指标之间的分数具有可加性，还需要对正态分数进行标准化。可以用下列过程描述：

$$z_{ij} = \frac{K(X_{ij} - \bar{X}_j)}{S_j} + C \tag{13.5.1}$$

其中， $z_{ij}$  是第  $i$  个个体在第  $j$  项一级指标上的正态标准分值； $X_{ij}$  是第  $i$  个个体在第  $j$  项一级指标上的正态分数。 $\bar{X}_j$  是常模中所有个体在第  $j$  项一级指标上的平均正态分值，即：

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \tag{13.5.2}$$

这里  $n$  为常模中个体的总数。 $S_j$  是第  $j$  项一级指标的标准差，即：

$$S_j = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n}} \tag{13.5.3}$$

$K$  和  $C$  只是调节常数，没有实际的意义，主要是为了防止出现可能的负数和保持转化后的标准分在形式上与原有分的相似。因此，在理论上  $K(K \neq 0)$  和  $C$  是可以任意设定的。

和原始分值相比较，标准分值具有以下一些特点：(1) 通过正态化标准分值，不但明确的表示了水平的高低，而且还表示出了个体在整个群体中的位置。(2) 采用正态化的标准分值，可以避免由于测试标准的不一致而带来的不可比性，使得个体间的比较成为可能。(3) 采用正态化标准分值，克服了不同指标之间的差异，使得计算个体的综合标准分值成为可能。个体的综合标准分值的计算类似于各个不同指标的正态化标准分值的计算，过程几乎一致。

图 13.5.1 描述了正态化标准分值与原始分值之间的差异。

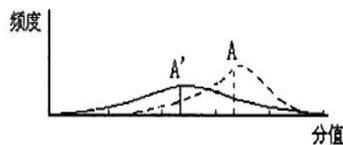


图 13.5.1 原始分与正态化标准分示意

在上图中，横坐标代表分值，纵坐标代表取得某个分值的个体数量。虚线是原始分的分布情况，实线是正态化标准处理后的得分分布情况。其中， $A$  点表示某个原始分值，而  $A'$  表示经过正态化以后的标准分值。有了常模及其正态化标准分值分布，就可以对特定个人进行测量和评价了。

当指标体系为层次结构时，首先，参照前面提出的各种量表对个体在各个下一级指标上的得分进行统计并计算出个体在该一级指标上的得分情况。再将个体的得分进行正态化处理，得到经过处

理后的百分等级分和正态分值,根据正态分值计算标准分值;然后,将各个下一级指标的正态化标准分值作进一步的综合处理,即得到上一级指标的正态化标准综合分,并最终得到该个体所具有的总体正态化标准综合分,以及在所有(或某个领域内)个体中所处的位置。

比如,若我们想从素质、能力和知识等方面综合评估一个人处理特定信息[解决特定问题]的能力,则解决特定问题[处理特定信息]的能力的综合评估方法可表达为:

$$H_s(A) = \sum p_{si} I_i(A) + \sum q_{sj} J_j(A) + \sum r_{sk} K_k(A) \quad (13.5.4)$$

其中,  $I_i$ ---主体 A 的特定素质 i 的评价值;  $p_{si}$ ---特定素质 i 对 s 类信息处理能力的影响系数;  $J_j$ ---主体 A 的特定能力 j 的评价值;  $q_{sj}$ ---特定能力 j 对 s 类信息处理能力的影响系数;  $K_k$ ---主体 A 的特定知识 k 的评价值;  $r_{sk}$ ---特定知识 k 对 s 类信息处理能力的影响系数。

其中,特定的素质、能力和知识包括分析问题和解决问题过程中的各类心理素质,与分析问题、解决问题有关的各种信息辨识、加工及操作能力,与解决问题有关的各种知识等。

### 3. 评估的基本原则

对心理素质、智力与能力进行综合评估是对人的评估,其评估应掌握以下原则:

① **系统性原则**。智力和能力具有系统的特点,智力和能力的形成是各种心理要素相互影响,形成非线性作用的结果。所以,对智力和能力等的评价不能只考虑单个因素,必须对各指标要素之间的关系进行全面考虑,分层次考虑各指标定性与定量评估的结合。

② **科学性原则**。指标体系要能客观、正确地反映要评估的智力和能力,要反映其科学的内涵,同时每一指标都应该科学地反映评估系统整体的某一个维度。指标设置要科学合理,有相对独立性,权重系数的确定要能正确反映各指标间的相互关系,以及各指标在总体评价中的地位和作用。要根据智力和能力的特点与实际,分析各种智力和能力的特征,正确划分指标层次。同一层次的指标选取,力争做到相互独立,并覆盖该层的整体情况。指标划分不重不漏,保证评价的科学性和合理性。

③ **可比性原则**。指标的设置数量要适当,指标的确定要尽可能明确、简单,便于操作。所列指标要素要尽可能是可量化的,而且对评价指标的赋值要有区分度,以便能用评价分值来区分不同能力要素的评分标准。个体间的评价结果要具有较好的离散性,便于横向比较。同时,评价中对于智力和能力只是一种量化的表述方式,避免因量化而量化,达不到评价目的。每个评估指标必须具有对于多个评估对象彼此可进行比较的特性,对不同的个体的智力和能力,其评分值要有显著性差异。

④ **协同性原则**。在建立指标体系的过程中,要力争做到指标之间的相关协同性。只有在各指标均衡协同前提下,才能保证整体能力水平的最优,满足非线性合成原则。从系统视角分析,智力和能力的构成要素相互作用、相互制约,它的整体功能决不是各要素的简单叠加,而是互相作用、相互制约中发生的协同作用,并呈现出一种非线性合成关系。因此,这些要素指标的不同组合方式将产生不同的能力绩效。科学的评价导向应有利于促进智力发展的效果,促进个体能力的形成。

⑤ **发展性原则**。智力和能力具有动态性、发展性特点,是随着时间的变化不断地发生变化的。若不学习,一个人的动态知识结构会随着时间的变化,思想和知识就有可能变得陈旧,能力也就趋于下降趋势。只有不断地学习,才能提高个人的知识和能力。因此,设计指标时要用动态性观点考虑这种学习与发展潜力,引导个体注意自己的潜力的可持续发展与不断动态提高。

### 3. 评估的检验

在测量学中,通常采用信度(Reliability)和效度(validity)两个指标对评估测量数据的真实性、可靠性以及测量的有效性进行检验。信度是测量的基本要素之一,缺乏信度的测量就不具意义,也

不可信。参照美国教育与心理测验标准中的相关定义,信度指的是测验分数未受测量误差(errors of measurement)影响的程度。简单的说,信度是反映了测量中的可靠性,也就是测量误差越小,测量结果就越可靠,信度也就越高。换言之,如果测量的结果能反应受试者真实的特征,而不因其它因素(如测验的环境、被检测个体的内部状态、测验题目的性质等)而影响其测验分数,那么这个测验所测量的结果是可靠的。误差是测量中常常不可避免的。在测量中最容易出现两种误差,一种是系统误差,还有一种是随机误差。一般而言,系统误差通常是集中的,向着一个方向分布的,在数量上表现为是恒常的;而随机误差则是由于种种随机原因造成的,很难加以控制,散布的方向不同,数量也是随机分布,很难预计。这是因为系统误差相对容易发现,并且容易克服,而随机误差则是由于一些不明原因造成的,很难发现也很难克服。

信度可以通过相关指标的估计加以判断。比如,可参照斯坦福大学柯隆巴克教授所发展的  $\alpha$  系数,依一定公式估量测验的内部一致性,作为信度估计的指标。其公式如下:

$$\alpha = \frac{k}{k-1} \left( 1 - \sum \frac{s_i^2}{s^2} \right) \quad (13.5.5)$$

其中,  $k$  是一级指标的数目;  $S^2$  是正态化综合标准得分的标准差;  $S_i^2$  是对是第  $i$  个指标的正态化标准得分中各类误差的标准差。这个公式表明:在  $k$  为常数的时候,  $\sum \frac{s_i^2}{s^2}$  越小,信度系数就越大,可靠性就越高。每个指标描述了要测量的一个领域,在每个领域中标准差应该比较小,而整个测量的领域之间应该有较大的差异,这样  $S_i^2$  就会很小,而  $S^2$  则会较大,从而保证整个测量的信度具有较高的值。也就是说,发散性误差较少,而系统性误差可以通过调整测量工具来达到减少的目的。根据柯隆巴克的理论,信度一般在 0.7 以上的测量可以认为是可信,可以作为个体衡量自身水平的一个重要参考依据。

衡量评价及其效果的另一个指标是效度(validity)。所谓效度是指根据测量结果推论变量特征的适合性(appropriateness),也就是测评的有用程度,以便考察所测的结果是否为测量者希望得到的。类似于信度系数的公式,效度系数一般规定为与测量的目的相关的分数的标准差在总标准差中所占的比例。美国社会学家巴比克在其《社会研究方法》一书中,对测量进行了深入的分析,并提出了效度的简单计算方式。如果假设  $X$  是实际测量得到的分数,那么,

$$X = T + B + E \quad (13.5.6)$$

其中,  $T$  表示真实值,  $B$  表示偏差即系统误差,  $E$  表示测量误差即随机误差。由于系统误差很难分解,因此可以简单的把系统误差包括到真实值之中。这是因为系统误差一般是同质的,和测量方法有关。在此条件下,实际得分可表达为:  $X = T + E$ 。

对于测量误差  $E$ , 由于其方向上的随机性,可以假设其期望值为 0, 并且与真实得分相独立。可以证明:  $E(X) = E(T)$ , 即实得分数和真分数的总体均值相等。同样可以证明:

$$s^2 = s_t^2 + s_e^2 = s_{\alpha}^2 + s_{\epsilon 0}^2 + s_e^2,$$

即实际得分的标准差等于真实分数的标准差与误差的标准差之和。其中,  $s_{\alpha}^2$  是真实得分的标准差,  $s_{\epsilon 0}^2$  是系统误差的标准差,  $s_e^2$  是随机误差的标准差。这样,可以定义效度系数的计算公式为:

$$\beta = \frac{s_{\alpha}^2}{s^2} = 1 - \frac{s_{\epsilon 0}^2 + s_e^2}{s^2} \quad (13.5.7)$$

该系数可用于描述测量中所采用的工具及其测量结果是否令人满意。

一般而言，分别有了信度和效度作为衡量心理测量工具及其测量结果的评价指标，即可根据这两个指标判断测量工具及其测量结果是否能令人满意了。

关于信度和效度之间的关系。我们可以从数理关系对信度和效度进行简单分析。首先，由于信度系数中的  $s_e^2$  不仅包含了真实分数的标准差，还包含了系统误差和随机误差的标准差，所以信度系数较大并不能保证  $s_x^2$  也一定较大，也就无法保证效度系数一定较大。其次，如果效度系数较大，

则  $s_x^2$  也较大，这样就说明无论是系统性误差还是随机性误差均较小，所以信度系数也一定较大。

由此，信度高是效度高的必要条件，而非充分条件；效度高则是信度高的充分条件，但不是必要条件。我们可以简单总结为：(1) 信度低，效度不可能高。如果测量的数据不准确，那么同样不能有效的反映被测对象的实际效果。(2) 信度高，效度未必高。(3) 效度低，信度未必低。(4) 效度高，信度必然高。

### 13.5.2 心理素质、智力与能力评估的模糊综合评估方法

#### 1. 模糊综合评估原理

由于智力、能力和心理素质等的抽象性特点，反映一个人智力、能力和心理素质等的指标系统，其中间差异不具有非此即彼的特征，比较适合运用模糊评估的方法。模糊评估方法的基础是扎德(L. A. Zadeh)教授所提出的用模糊集理论描述事物的方法。模糊集合理论认为，一个对象  $x$  属于某个集合  $A$  的程度可以是 0 或 1，还可以是某个中间值。即其对一个集合  $A$  的隶属程度可以用“隶属函数”  $\mu_A(x): 0 \leq \mu_A(x) \leq 1$  来表征。这里， $A$  为“模糊集合”， $\mu_A(x)$  为对象  $x$  对  $A$  的“隶属度”。

模糊评价法在对被测对象进行评价时，是根据不同的工作或任务要求，依据不同的测评目的，对个体智力活动过程的一些行为和心理活动进行模糊处理，运用模糊综合评估原理进行测评。

假设有两个有限论域集合： $U=[u_1, u_2, \dots, u_n]$ ， $V=[v_1, v_2, \dots, v_n]$ ；其中，假定  $U$  代表评价要素集合， $V$  代表评价结果集合。模糊评估为：

$$X \circ \tilde{R} = Y \quad (13.5.8)$$

式中  $X$  是  $U$  上的模糊子集，代表评价要素； $Y$  是  $V$  上的模糊子集，代表评价结果； $\tilde{R}$  是评价矩阵。智力和能力的特性决定了智力和能力的评估具有复杂性与模糊性的特点，很适合于用此法给出模糊量化结果。

层次分析法 (Analytic Hierarchy Process AHP) 是萨迪 (T. L. Saaty) 提出的，是从定性到定量分析综合集成的系统工程方法。根据具有递阶结构的目标、子目标、约束条件等用两两比较法确定判断矩阵，然后把矩阵最大特征根所属特征向量的分量作为系数，综合出各方案权重。此法基于以下原理：

给定有限集  $S, T$ ；这里， $S$  是独立属性集合， $T$  是对象集合，假设： $\forall s_i \in S, \exists$  (存在) 相对重要性指数  $w_j > 0, j=1, 2, \dots, n$  并且  $\sum w_j = 1$  假设  $\forall t_i \in T, \exists$  (存在) 相对于  $s_j$  的“权”  $w_{ij} > 0, i=1, 2, \dots, m$ ，并且  $\sum w_{ij} = 1$  则，诸  $w_{ij}$  的凸组合表示  $t_i$  对于  $S$  的相对重要性程度。

### 2. 层次指标体系的模糊综合评判

作为一个示例，我们考虑层次指标体系的模糊综合评估问题。对人类个体进行评估的指标体系是多种多样的，依据不同的评估侧重可建立不同的指标体系，而且这些指标体系也需要根据测量对象、人类本身的发展等情况的变化而不断的修正和完善。

设有一个多因素的递阶层次指标体系，其结构图如图 13.5.2 所示。

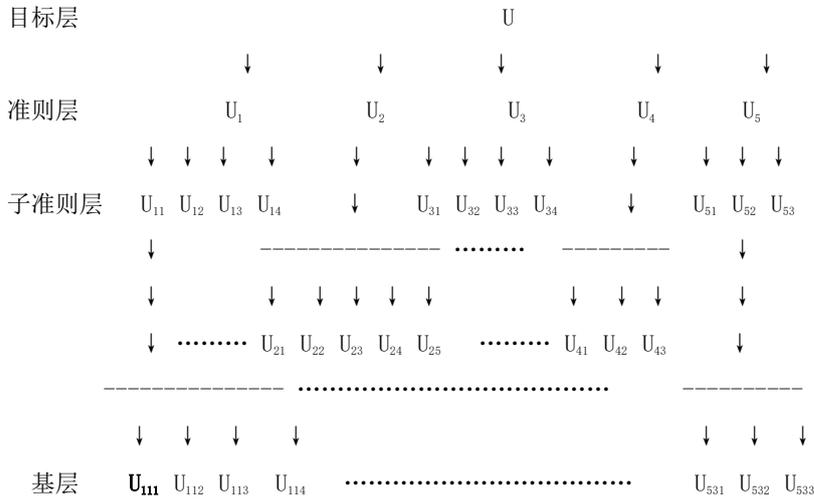


图 13.5.2 心理评估层次指标体系

有了初步确定的指标体系，就可以建立必要的评判方法。在这里，采用的是层饮分析和模糊综合评判相结合的方法。

首先，确定模糊集合： $U = \{U_1, U_2, U_3, U_4, U_5\}$ ；其中， $U_1 = \{U_{11}, U_{12}, U_{13}, U_{14}\}$ ； $U_2 = \{U_{21}, U_{22}, U_{23}, U_{24}, U_{25}\}$ ； $U_3 = \{U_{31}, U_{32}, U_{33}, U_{34}\}$ ； $U_4 = \{U_{41}, U_{42}, U_{43}\}$ ； $U_5 = \{U_{51}, U_{52}, U_{53}\}$ 。 $U_{11} = \{U_{111}, U_{112}, U_{113}, U_{114}\}$ ； $\dots$ ； $U_{53} = \{U_{531}, U_{532}, U_{533}\}$ 。再确定评语集合。设  $V = \{v_1, v_2, v_3, v_4, v_5\}$ ，即评语采用 5 级评分，集中的 5 个元素分别对应评语“优、良、中、可、差”。

其次，采用层次分析法(AHP)确定各指标的权重。层次分析法可以量化评估者的经验判断，在目标结构复杂且缺乏必要数据的情况下比较实用。步骤是先画出多因素递阶层次结构图，然后用专家咨询法构造两两比较判断矩阵，再用方根法或和积法求矩阵的特征向量和特征根，具有满意一致性的评价矩阵对应的特征向量各分量即为各指标对上层指标的权重。设， $U_{ijl}$ 对  $U_{ij}$ 的权重为  $a_{ijlk}$ ，且  $A_{ij} = (a_{ij1}, a_{ij2}, \dots, a_{ijl})$ 。其中  $l$ 为基层层中指标的个数。 $U_{ij}$ 对  $U_i$ 的权重为  $a_{ij}$ ，且  $A_i = (a_{i1}, a_{i2}, \dots, a_{ij})$ 。其中  $j$ 为子准则层中指标的个数。 $U_i$ 对  $U$ 的权重为  $a_i$ ，且  $A = (a_1, a_2, a_3, \dots, a_i)$ 。其中  $i$ 为准则层中指标的个数。

确定评价隶属度矩阵  $R_{ij}$ 。可采用单因素模糊评价法，为每个因素确定评价对象对评价集各元素的隶属度。设  $U_{ijl}$ 中第 1 个元素对  $V$ 中第  $k$  元素的隶属度为  $r_{ijlk}$ 。则  $U_{ijl}$ 的评价结果是  $R_{ijl} = \{r_{ijl1}, r_{ijl2}, r_{ijl3}, r_{ijl4}, r_{ijl5}\}$ 。其中， $r_{ijlk} = v_{ijlk} / \sum v_{ijlk}$  ( $v_{ijlk}$ 是参评专家中认为指标  $U_{ijl}$ 属于  $v_k$ 等级的专家人数)评价隶属度矩阵为：

$$R_{ij} = \begin{bmatrix} r_{ij1} \\ r_{ij2} \\ \vdots \\ r_{ijl} \end{bmatrix} = \begin{bmatrix} r_{ij11} & r_{ij12} & r_{ij13} & r_{ij14} & r_{ij15} \\ r_{ij21} & r_{ij22} & r_{ij23} & r_{ij24} & r_{ij25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{ijl1} & r_{ijl2} & r_{ijl3} & r_{ijl4} & r_{ijl5} \end{bmatrix} \quad (13.5.9)$$

至此,可以进行多因素模糊综合评判。先对各子准则指标的评价矩阵  $R_{ij}$  做模糊矩阵运算,得到子准则指标对于价集  $V$  的隶属向量  $B_{ij}$ :

$$\begin{aligned} B_{ij} &= A_{ij} \otimes R_{ij} = (a_{ij1} \quad a_{ij2} \quad \cdots \quad a_{ijl}) \begin{bmatrix} r_{ij11} & r_{ij12} & r_{ij13} & r_{ij14} & r_{ij15} \\ r_{ij21} & r_{ij22} & r_{ij23} & r_{ij24} & r_{ij25} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{ijl1} & r_{ijl2} & r_{ijl3} & r_{ijl4} & r_{ijl5} \end{bmatrix} \\ &= (b_{ij1} \quad b_{ij2} \quad b_{ij3} \quad b_{ij4} \quad b_{ij5}) \end{aligned} \quad (13.5.10)$$

其中:“ $\otimes$ ”是模糊矩阵合成算子符号。

$$R_i = \begin{bmatrix} l_{i1} \\ \vdots \\ l_{ij} \end{bmatrix} = \begin{bmatrix} b_{i11} & b_{i12} & b_{i13} & b_{i14} & b_{i15} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{ij1} & b_{ij2} & b_{ij3} & b_{ij4} & b_{ij5} \end{bmatrix}$$

再对子准则层指标的评价矩阵  $R_i$  进行模糊矩阵运算,得到准则层指标对于评价集  $v$  的隶属向量  $B_i$

$$B_i = (l_{i1} \quad a_{i2} \quad \cdots \quad a_{ij}) \otimes \begin{bmatrix} b_{i11} & b_{i12} & b_{i13} & b_{i14} & b_{i15} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{ij1} & b_{ij2} & b_{ij3} & b_{ij4} & b_{ij5} \end{bmatrix} = [b_{i1} \quad b_{i2} \quad b_{i3} \quad b_{i4} \quad b_{i5}] \quad (13.5.11)$$

$$R = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix} \quad (13.5.12)$$

对准则层指标的评价矩阵量进行模糊矩阵运算,得到目标层指标对评价集  $V$  的隶属向量  $B$ :

$$\begin{aligned} B &= A \otimes R = (a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5) \otimes \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{bmatrix} \\ &= (b_1 \quad b_2 \quad b_3 \quad b_4 \quad b_5) \end{aligned} \quad (13.5.13)$$

如果  $\sum_{k=1}^5 b_k \neq 1$  时,进行归一化处理,即  $\bar{b}_k = \frac{b_k}{\sum_{k=1}^5 b_k}$ , 这样可以得到:

$$\bar{B} = (\bar{b}_1 \quad \bar{b}_2 \quad \bar{b}_3 \quad \bar{b}_4 \quad \bar{b}_5) \quad \bar{B} \text{ 就是 } U \text{ 对 } V \text{ 的隶属向量,也就是评价结果。}$$

最后,对评价结果用加权平均法进行处理,这样将评价集  $V$  中的各个因素量化后,最终得到评价结果:  $V = B \bar{V}^T$ 。

这里,只是利用层次分析法和模糊综合评判,以非常简单的方式阐述了如何进行模糊综合评估。

关于层次分析法和模糊综合评判的具体计算过程不再赘述，可以参见相关参考文献。

### 13.5.3 对人才和能力的综合判定

社会对人才的需求是一种有针对性的需求。对于一个社会组织来说，其运行与发展需要各种人才，需要由具有不同业务能力和管理能力的各种人员来构成一个有机的整体，需要由具有不同智力因素、不同知识水平、不同能力层次的人才组合成合理的群体智能结构。因而，其对人才的需求、评估与测度也是针对不同类型的。不同的需求要求人才必须具备的素质结构、知识结构和能力结构各不相同。其中，一个人的素质包括生理素质、心理素质、文化素质和行为素质等。其素质结构是指其素质各个组成部分及其相互联系的方式，是素质内在要素的表现形式，是作为完整的特质表现出来的有机整体。所谓素质测评则是指一种建立在素质特征信息“测”与“量”基础上的分析与评判活动，即以量化的方式对人的能力倾向、个性特征进行测试的活动。其知识结构和能力结构是指为实现或达到既定目标，行为主体应具备的知识和能力的构成和内部结构，是个人所具有的知识、智力和能力所组成的多序列、多要素、多层次的动态的有机综合体，是由人的知识要素与能力要素相互连结、相互融合而形成的整体功能。不同的工作，所需人才的知识、技能、素质各异。也就是说，有个结构问题。一个人的知识和智能结构，是在学习、劳动和社会生活中逐渐形成的。不同的智能结构，造成了各种人才的特点、长处，形成了各种人才的不同类型特点。智能体现了一个人驾御各种资源的能力。因此，评价一个人的能力实际上就是通过一些相关的指标评判获得对其整体能力体系的评价。随着知识时代的发展，智能更多的体现在知识的积累和学习的能力的提高方面。相应的，信息的收集和利用能力、知识的获取和运用能力、决策与解决问题能力等也就成为衡量一个人的整体能力的关键组成部分。社会对人才的能力要求，主要包括：

(1) **信息感知、采集及分析能力。**从古至今，信息对于智能而言，其重要性一直不容置疑。在进入信息社会的今天，信息的价值更已为大家所接受。收集信息和分析信息的能力已经成为制约人的能力不断发展的重要因素。一般而言，信息的收集和分析能力实际上反应了人的信息素质。一方面，收集信息需要有相关的信息渠道和信息收集技术，尽管传统的人工调查和信息汇总也是可以考虑的方法，但是从投入的成本、收集到的信息数量和质量等方面进行衡量，显然要比采用现代的因特网和计算机技术方法进行信息收集和汇总要相差许多。另一方面，随着现代社会信息化程度的不断提高，信息的种类和数量急剧增加，形成了所谓的“信息爆炸”之说。在这样如此众多的信息中，有许多是有价值的，但是也有许多是无用的、甚至是有害的信息，如何进行分析处理就成为了人的一种素质或能力。所谓信息分析处理就是利用相关的技术和方法从数量巨大的信息海洋中去伪存真、深入挖掘，以获取有益的信息。因此，衡量当代人信息收集和分析能力的一个重要指标就是其可获得与利用信息技术的程度，特别是对信息进行深度挖掘的程度。

(2) **信息加工利用能力。**拥有信息并不一定就等于拥有了财富。收集信息、分析信息的目的是为了利用信息，让有价值的信息在适当的时间、适当的地点、交给适当的使用者（包括自己），使其能够根据这些有价值的信息进行恰当地处理，充分发挥这些信息的作用，带来实际的利益。事实上，无论是收集到的信息，还是根据收集到信息进行深度挖掘而得到的信息背后的信息，充其量只是客观事实的反应，只是表明了某种客观事实或者有可能会成立的事实；如何根据这些事实充分发挥决策者的聪明才智、充分的利用这些信息进行决策才能真正体现出信息的价值。

(3) **相关知识的掌握程度。**人的智能一个显著特征是尽可能的掌握知识。无论在什么样的工作中，知识都是必要的。知识和经验需要学习和掌握。在知识日益重要的今天，如何保证拥有不断丰富的

知识、保证这些知识不会随着知识的更新而过时，已经成为智能的重要因素。人才，都是掌握有丰富知识的人，都是善于学习的人。

(4) **相关知识的灵活运用程度**。就像即使知道烤面包的所有程序却仍然无法烤出好的面包一样，没有对知识的灵活运用，再多的知识也没用。有时，知道实际该怎么做比知道理论上该怎么做更重要。从理论上来说，智能的建设和提高包括着“知”和“行”两个环节。人才，是掌握知识的人，更是可灵活运用其所掌握的知识处理现实生活中的问题的人。

(5) **决策的准确与迅速程度**。决策不但可以充分发挥人的聪明才智，而且还可以通过知识的运用特性，反应出每个人的实践能力。一个人的决策水平，同时也就提供了其自身的知识、智力和能力的水平。

(6) **相互协调与协作的能力**。知识的掌握和运用能力，也即决策能力，无疑会使人的聪明才智得到充分发挥，但是，仅仅局限于此无疑也是狭隘的。人是社会的一分子，通常也是团体中的一分子。人才，只考虑其个体能力的发挥是远远不够的，还需要在其合作的各成员之间形成协调的协作关系。协调能力是保证组织成员的智慧能够不断融合、提升组织的共同思考、提升组织整体的决策能力和执行能力的重要条件。事实上，组织成员之间的相互协调、知识共享的能力是一种信息共享、精诚合作和相互信任的文化氛围，是实现组织的共同决策和共同思考的保证。实质上，它是组织内部的“大脑连线”，有了它，才使组织真正成为个体的延伸。因此，现代社会注重个人的能力，更注重其可相互协作的能力。

实际上，社会和组织对其所需人才的评定并没有一定之规。或注重能力，或注重素质，或注重知识，或注重品行。古人尚且不排除“鸡鸣狗盗之徒”，现代社会更需要各方面的人才。各组织对人才的不同要求，实际上所反映的是各个组织对各类素质或能力对其不同需求的作用，是根据其不同需要来对人才进行评定的。

对各类考察对象的各类素质或能力的测试方法有笔试、面试、观察、试用等。对人才的评定也可使用多种方法。比如，我们可使用**高标定位法**来对人才进行评定。其方法是，首先算出理想化的人才的标准化向量，然后将被考察对象的考察结果与这个标准化向量比较，从而确定出被评估个体的相对符合程度。评定时可采用相似系数法。根据高沃(Gower)提出的比较定性和定量变量的相似系数法，对于定性变量，该法是将有多个取值的变量先化为二态变量，然后计算其匹配程度。高沃提出的相似系数公式为：

$$G = \frac{\sum_{i=1}^m W_i s_i}{\sum_{i=1}^m W_i} \quad (13.5.14)$$

这里， $m$ 为变量个数， $W_i$ 表示权重，当两样本第 $i$ 个变量值存在时， $W_i=1$ ；当样本第 $i$ 个变量值不存在时， $W_i=0$ 。 $0 \leq s_i \leq 1$ ， $s_i$ 表示两样本在 $i$ 上的相似程度。这里，我们定义：

$$s_i = 1 - \frac{|x_i - x_{0i}|}{R_i} \quad (13.5.15)$$

这里， $R_i$ 表示 $i$ 在样本总体中的变化范围，对所有样本均计算与标准向量的相似系数 $s_i$ ， $s_i$ 越大，表示样本与标准向量越接近，而标准向量就是具有专业特征的理想人才。这种方法可以使

对欲选人才在多大程度上可满足所需特定类型人才的条件进行评价，也可判定其距离理想型人才的差距及其努力的目标与方向。

下面，我们以创新能力的模糊综合评估来作进一步地说明。

创新能力模糊综合评价过程如下。

### 1. 明确创新能力评估的基本假设

我们假设，决定创新能力的要素是一个相互关联的复杂系统，只能通过模糊综合评估法来确定其量化指标。所建构的模型已是从多视角进行了分析，符合创新能力评估的要求与发展潜能预测的要求。在评估系统内，可根据 AHP 法建立创新能力评价指标体系层次结构模型。模型将有关各要素按不同属性自上而下分成多层，同一层次诸要素只从属于上层要素，同时又支配下层因素。上层是能力评价目标，中间层是分解的几个维度，下层是具体的创新能力评价指标体系。

### 2. 进行创新能力评估的要素分析

由于创新能力是多种素质和能力综合作用的结果，可以从多个角度对创新能力进行评估。创新能力评估的指标体系设计，首先要确定从哪些方面来考察创新能力。我们可给出创新能力的一套评估维度，如：思维能力、想象能力、信息观察采集能力、信息综合与转化能力、学习能力等。认为这几个维度可构成创新能力的评价指标体系。

### 3. 创新能力评估的指标系统设计

首先，可参考多种公认的评价标准，即判定原则，从中归纳产生出创新能力系统各要素的初评指标。然后，使用德尔菲法通过与专家的问卷与访谈，最终确定各指标内容。

一般而言，创新能力评价指标体系可采用频度统计、理论分析、德尔菲专家访谈等办法来设置、筛选。在此基础上，可应用独立性分析、系统要素协同分析等方法，选择出内涵丰富又相对独立的指标构成评价指标。整个评价指标体系设计过程可如图 13.5.3 所示。

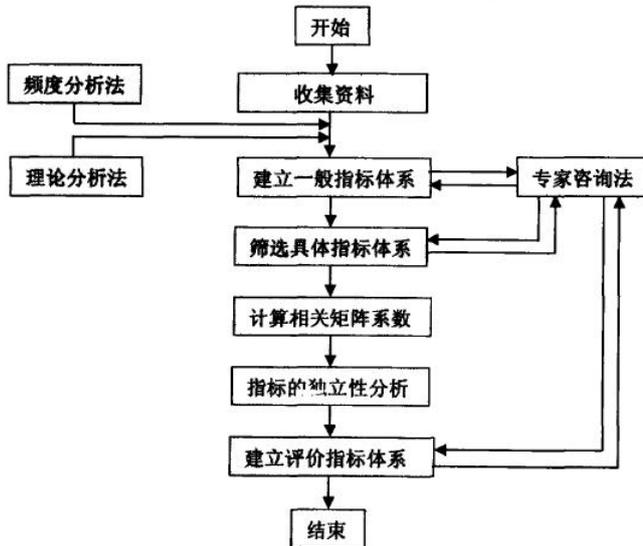


图 13.5.3 创新能力评价指标体系设计过程

### 4. 建立创新能力评估的数学模型

其方法可如下：

#### (1) 设定各因素综合权重

创新能力评价为评估体系  $U$  上的模糊集  $X$ ，设各因素权重为：

$$A = \{A_1, A_2, \dots, A_M\};$$

$$A_i = \{A_{i1}, A_{i2}, \dots, A_{iN}\}; A_{ij} = \{A_{ij1}, A_{ij2}, \dots, A_{ijR}\}; \dots\dots\dots$$

并满足  $\sum A_i = 1, \sum A_{ij} = 1, A_i \geq 0, A_{ij} \geq 0, i=1, 2, \dots, M; j=1, 2, \dots, N$ 。

(2) 运用改进的模糊层次分析方法, 建立模糊评判矩阵

改进的模糊 AHP 法的程序系统结构图如图 13. 5. 4 所示。

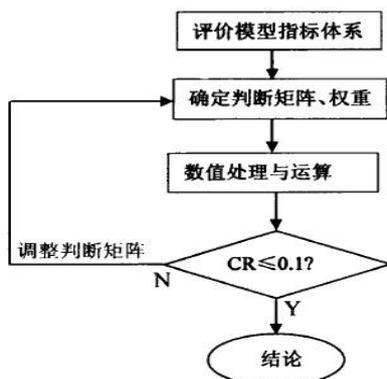


图 13. 5. 4 改进的模糊 AHP 法程序系统结构图

运用 AHP 法与模糊数学法综合评定需要确定各指标权重，评价因素的权重是每个因素在集合中的重要程度的定量表示，即所占的比重，其合理性将影响评价结果的合理性。权重的确定可以采取德尔菲法，对选定专家组进行德尔菲调查问卷，根据调查结果确定各指标值，并按归一化要求对各因素赋予相应的权重。

运用改进的层次分析方法，确定模糊综合评判的权重可首先设定评判集： $P = \{P_1, P_2, \dots, P_k\}$ ，这里， $P_1, P_2, \dots, P_k$  分别指很重要、较重要、一般、不重要等。然后让专家评定，所使用的专家调查表可如表 13. 5. 1 所示。

表 13. 5. 1 创新能力指标要素合理性专家评价表

评估等级 指标因素	$P_1$	$P_2$	$P_3$	$P_4$	$P_5$
$U_1$					
$U_2$					
...					
$U_m$					

根据评价意见，可综合各方意见确定归一化后的权重向量  $\tilde{a}_i$ ； $\tilde{a}_i$  是指标要素子集  $u_i$  的权重集合。

(3) 建立模糊评价集

评价因素可以分为定性因素和定量因素，两种因素要分别进行分析处理。其中，定性因素的模糊综合评价可以通过模糊统计方法求得。即对需要评估的因素设定一个评价集  $V = \{v_1, v_2, \dots, v_n\}$ ，这里  $v_j (j=1, 2, \dots, n)$  表示对评估因素进行评价时的第  $j$  个等级。如很好、较好、一般、较差、很差等。然后让评估专家分别对各被评估对象在各个评估指标上的评估值进行打分。由此，我们即可采用专家评判法得到单因素评价矩阵  $\tilde{R} = (\tilde{R}_{ij})_{m \times n}$ ， $\tilde{R}_{ij}$  表示被评估对象在指标  $X_i$  方面被评为  $v_j$  的模

模糊隶属度。定量因素的模糊综合评价则可通过建立隶属函数来确定。

(4) 逐次模糊综合

利用公式  $\tilde{b}_i = \tilde{a}_i \circ \tilde{R}_i$ ，可求得第一层次各因素的模糊综合评判集合，这其中  $\tilde{a}_i$  是指标要素子集  $u_i$  的权重集合。为了统筹兼顾各种因素的影响，可利用“加权求和型”广义模糊算子  $M(\cdot, \oplus)$  来计算

$$\begin{aligned} \tilde{b}_i &= \tilde{a}_i \circ \tilde{R}_i \\ &= b_{ij} = (\mu_{u_i}(v_1), \mu_{u_i}(v_2), \dots, \mu_{u_i}(v_3)) \\ \text{其中, } i &= 1, 2, \dots, n: \\ \mu_{u_i}(v_k) &= \min(1, \sum_{j=1}^{p_i} a_{ij} \mu_{ij}(v_k)) \\ \text{其中, } i &= 1, 2, \dots, n; \quad k = 1, 2, \dots, 5; \end{aligned} \tag{13.5.16}$$

这时的权重集和创新能力评价模糊集都满足归一化条件。

在一级模糊评判的基础上，可以求得上一层次直至第一层次要素集合  $U = \{u_1, u_2, \dots, u_n\}$  上的模糊综合评价矩阵  $\tilde{R} = [\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_n]^T$  以及创新能力程度的模糊综合评判集合  $\tilde{B} = \tilde{A} \circ \tilde{R}$ ，其中  $A$  是指标要素集  $U$  的权重集。

(5) 模糊综合评价

为了综合比较，有时需要求出创新能力的分数。创新能力的分数可综合评定为：

$$E = E(\tilde{B}) = \tilde{B} \cdot H = \sum_{i=1}^n \mu(v_i) h(i) \tag{13.5.17}$$

其中， $B$  为创新能力的模糊综合评判向量， $H$  为设定的对模糊评价的数值化向量。

### 13.6 人类智能测评未来的发展趋势

智力和能力测验虽历经百年，并取得了长足的进步，但与许多成熟领域相比还只能算是刚刚起步。随着人们对人类智能认识的不断深入，智力和能力的测验还会进一步持续发展。纵观当前对智能及智能测评研究的发展取向，可以预见，未来智能测评的演变将主要表现在以下几个方面。

第一，越来越多的研究者认识到元认知能力是智力和能力的重要组成部分。所谓元认知是主体对自身认知活动的认知，对个体的基本认知活动起着重要的计划、监测和调控作用，并因此影响着基本认知活动的过程和效果。比如，戴斯在其智力的“计划—注意—同时性加工和继时性加工”模型中强调，负责计划、监控、调节、评价等元认知功能的计划系统是智力系统的重要组成部分，并在其编制的 CAS 测验中专门设计了计划测验来测量被试的元认知能力。也有研究者提出，智力包含三种水平：能力、元认知和目的性。因此，可以预见，随着人们对元认知在智力系统中作用认识的深入，在智力测验中增加对元认知的测量将会成为一个新的发展方向。

第二，传统智力测验对个体创造力的评估有限，而解决新奇问题的创造力作为智力的一部分已经为许多心理学家所认可。有研究者强调，创造性是智力的三个重要组成部分之一。斯滕伯格的成功智能理论通过大量实验也证实，成功智力由分析性、实践性、创造性三个成分平衡而得，而传统智力测验只是测量了其中的分析性成分。因此他专门设计了成功智能测验(即斯滕伯格三元能力测验，简称 STAT)分别评价分析性、实践性和创造性三个方面，并通过大量实验验证了 STAT 的信度和效度。

而戴斯等人经过大样本和个案研究也证实,其 PASS 理论和 CAS 测验比传统智力测验能更好地预测创造力、鉴别天才儿童。STAT 和 CAS 测验为未来智力测验的编制已经进行了非常有价值的开拓性工作。今后,从心理学的本质上厘清创造力和智力的关系、开发更有效预测创造力的智力测验,将会成为一些研究者的重要目标。

第三,智力测验最初的目的是甄别出那些需要接受特殊教育的儿童,若仅从这个目标出发,应该说大多数智力测验基本上都胜任了这一任务,并且相当一部分智力测验的结果也可以大致预测某些特殊个体未来智力的发展。然而,为儿童的教育和干预服务更应该成为智力测验的最终目标之一,从这个角度来看,传统智力测验在诊断、预测和干预等方面还存在很多不足。因此,未来的智力测验应当对被试作出更为具体、准确以及有针对性的评价,并为有效的干预措施提供必要的信息。对研究者来说,这意味着一方面要从理论上明确智力的性质和具体成分,以此来指导智力测验的编制,另一方面,要依据有关的智力理论确定具有代表性的测验项目,并明确各测验项目的认知含量以及权重,特别是要以智力测验所作出的诊断为基础,提出相应的干预措施,为儿童以及智力发展落后者提供必要的指导和训练,进而提高其智力水平。在这方面,基于认知过程的PASS理论和CAS作出了有益的尝试,在鉴定和准确地解释被试智力迟滞原因等方面均优于传统的智力测验,而且还提出了相应的PASS补救计划(PASS Remediation Program,简称PREP),这对未来智力测验的研究和编制无疑是一个很好的启迪。

目前,对智力的测验,总的发展趋势是:(1)智力测验的内容:智力理论的发展正促使智力测验的内容多元化;(2)智力测验的编制:分析技术的进步正推动智力测验编制的科学化;(3)智力测验的实施:由静态测试转向动态评估;(4)智力测验的效用:由鉴别、筛选和预测正转向矫正、发展和提高。而对成人能力的测验,也正在逐步科学化、规范化、系统化和多样化。

#### 主要参考文献

- 1301 赵笑梅 陈英和 智力个体差异研究述评 心理科学进展 2007, 15 (3): 451-457
- 1302 钟建军 智商常态人群的智力个体差异研究述评 心理科学 2007, 30(2):394-399
- 1303 王玲 沈政 樊春雷 智力与心理速度的关系研究述评 心理科学进展 2006, 14 (1): 40-45
- 1304 马妍妍 李寿欣 个体智力差异的脑电生理学基础—来自 EEG 和 ERP 的证据 心理科学进展 2007, 15 (6): 872-877
- 1305 陈英和,赵笑梅 智力测验的演变和展望 北京师范大学学报(社会科学版) 2007 年第 3 期
- 1306 姚伟 华道金 多元智能理论的评价观及其对我国幼儿发展评价的启示 外国教育研究 2004 年第 9 期
- 1307 BJORKLUND D F. Children' s Thinking [ M ] . Belmont : Wadsworth/ Thomson Learning , 2003.
- 1308 (美) A · 安娜斯塔西, S · 厄比娜. 心理测验[M] 缪小春,竺培梁译. 杭州:浙江教育出版社, 2001.
- 1309 STERNBERG R J , KAUFMAN J C. Human abilities [ J ] . Annual Review of Psychology, 1998 , (49): 479-502.
- 1310 STERNBERG R J. A broad view of intelligence: the theory of successful intelligence [J]. Consulting Psychology Journal: Practice and Research, 2003, 55 (3):139-154.
- 1311 STERNBERG R J. Successful intelligence: finding a balance[J]. Trends in Cognitive

Science, 1999, 11 (3):436-442.

1312 BEIER M E, ACKERMAN P L. Current -event s knowledge in adult s : an investigation of age , intelligence , and nonability determinants[J ] . Psychology and Aging, 2001 , 16 (4):615 - 628.

1313 REEVE C L, HA KEL M L. Toward an understanding of adult intellectual development: investigating within individual convergence of interest and knowledge profiles[J ] . Journal of Applied Psychology, 2000, 85 (6): 897 - 908.

1314 STERNBERG R J. A t riarchic approach to the understanding and assessment of intelligence in multicultural populations[J] . Journal of School Psychology, 1999, 37(2) :145 - 159.

1315 SHAVININA L V. Beyond IQ: a new perspective on the psychological assessment of intellectual abilities[J]. New Ideas in Psychology , 2001 (19):27 - 47.

1316 GRIGOREN KO E L, STERNBERG R J. Dynamic testing [J] . Psychological Bulletin , 1998 , (124) : 75 -111.

1317 STERNBERG R J . The concept of intelligence and it s role in lifelong learning and success[J] . American Psychologist, 1997, 52 (10) :1030 - 1037.

1318 EMBRETON S E. The new rules of measurement[J]. Psychological Assessment, 1996, (8) :341 - 349.

1319 OKLAND T, HAMBLETON R K. International Perspective on Academic Assessment [M] . Boston : Kluwer Academic Publishers , 1995.

1320 潘菽. 人类的智能. 上海: 上海科学技术出版社, 1985

1321 皮连生. 论智力的知识观. 华东师范大学学报(教育科学版), 1997, (03).

1322 吴天敏. 关于智力的本质. 心理学报, 1980, (03).

1323 Bandura A, Locke. Negative self-efficacy and goals revisited . Journal of Applied Psychology, 2003, 88 (1) :87-99

1324 Caspi, & Silva. Temperamental qualities at age three predict personality traits in young adulthood: Longitudinal Evidence from a Birth Cohort . Child Development, 1995, 66(2) :486-498

1325 Harackiewicz J M, Elliot A J. Achievement goals and intrinsic motivation . Journal of Personality and Social Psychology. 1993, 65 :904-915 .

1326 Utman C H. Performance effects of motivational state:A meta-analysis . Personality and Social Psychology Review, 1997, 1 (2) :170-18

1327 Vansteenkiste M, Deci E L. Competitively contingent rewards and intrinsic motivation: Can losers remain motivated . Motivation and Emotion, 2003, 27 (4) :273-299

1328 胡瑜, 孔克勤. Personality:人格还是个性. 心理科学, 2006, (2)

1329 张春兴. 现代心理学. 上海: 上海人民出版社, 1994

1330 张野, 杨丽珠. 我国 3~6 岁儿童个性类型及发展特点的研究. 心理科学, 2005, (4)

1331 郭永玉, 张钊. 人格心理学的学科架构初探. 心理科学进展, 2007, (2)

1332 黄坚厚. 人格心理学. 台北: 心理出版社, 1999

1333 王登峰, 崔红. 中西方人格结构差异的理论与实证分析——以中国人人格量表(QZPS)和西方五因素人格量表(NEOPI-R)为例. 心理学报, 2008, (3)

1334 张建新, 周明洁. 中国人人格结构探索——人格特质六因素假说. 心理科学进展, 2006, (04)