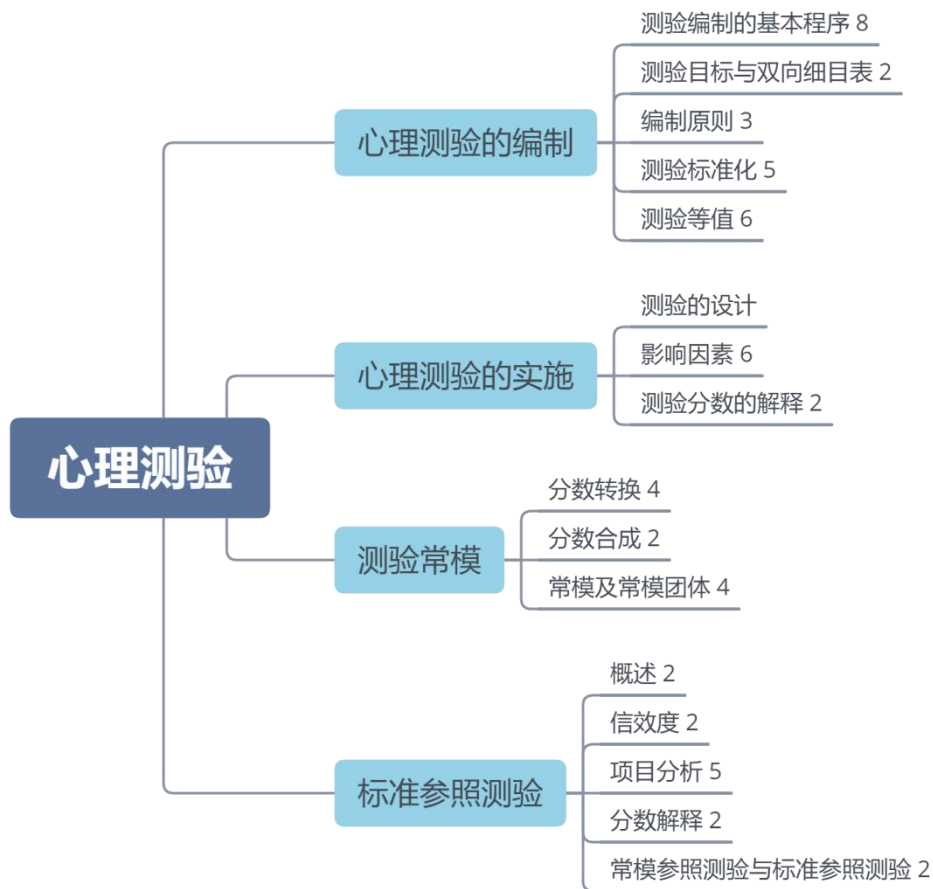


第四章 心理测验



一、心理测验的编制

(一) 测验编制的基本程序

1. 确定测验目的：明确测量对象、确定测量目标、明确测量用途（描述性、诊断性）。
2. 制定编题计划：明确测验的内容，使内容全面而具有代表性；明确各个内容的相对重要性。
3. 编辑测验项目（题目编制技术）：收集测验资料（丰富、普遍、有趣味），选择测验项目形式（简答题、论文题等），编写测验项目（对照双向细目表）。
4. 预测与项目分析：预测对象有代表性，预测情境要和正式施测时一致；对项目的难度、区分度等进行分析。
5. 合成测验：完成测试项目的选择、编排（并列直进式、混合螺旋式）及复本的编制。
 并列直进式是将整个测验分为若干分测验，属于同一分测验的项目，依其难度由易到难排列；混合螺旋式则是先将各类项目按难度分成若干不同层次，再将不同性质的项目予以组合，作交叉式的排列，难度渐次上升，其优点是被试对各类项目循序作答，从而维持作答的兴趣。
6. 测验的标准化：测验的编制、施测、评分及解释测验分数的程序的一致性。

7.鉴定测验：主要是鉴定测验的信度、效度、测验量表及常模。

8.编写测验说明书：向使用者说明测验目的及功用、理论依据以及项目选择依据、实施方法、测验的信效度、测验适用人群的特点和性质、测验的答案及评分、常模表等。

（二）测验目标与双向细目表

1.测验目标

测验目标是指测验是用来测量什么心理变量或行为特征的，需有操作定义。

2.命题双向细目表（命题蓝图）

命题双向细目表是测验编制过程的重要依据，相当于编题的计划清单，包括两个维度的表格（左一列表示测验内容，第一行表示测量结果，数据表示某内容的权重）。它指出了测验所包含的内容与要测定的各种技能，以及每个内容和技能的相对重要程度。

双向细目表可以使命题工作避免盲目性，而具有计划性；使命题者明确测验的目标，把握试题的比重和分量，提高命题的效率和质量；可提高测验的内容效度，克服专家评定法的一些不足。

（三）编制原则

1.测验编制的基本原则

信度高、效度高、难度适中、区分度强。

2.题目编制的一般原则（技术）

- （1）项目取样有代表性；
- （2）项目取样范围同编题计划所列项目范围一致；
- （3）初编题目数量多于最终所需数量（一般是多一倍），便于筛选；
- （4）题目用语要求简练、明了；
- （5）题目格式不被误解；
- （6）没有引起争议的答案；
- （7）避免涉及社会禁忌和个人隐私；
- （8）题目间相互独立；
- （9）项目难度有一定的分布范围；
- （10）力求施测和评分省时省力。

3.测题的种类及编制

根据受测者的应答方式，题目分为客观题（固定应答型题目）和主观题（自由应答型题目）。

（1）客观题

①是非题：从是非两个答案中做出选择。

A.优点：出题容易、回答方便、评分简单、使用广泛。

B.缺点：只适用于考查对简单观念或知识的了解，容易造成猜测。

②选择题：从选项中选择一项自认为正确的答案。

A.优点：不仅评分简单、省时、客观，而且相比是非题，它更少受猜测的影响，使用广泛。

B.缺点：编拟迷惑答案比较困难，无法测量出被试的言语表达能力和概括、组织能力，

猜测影响仍然不能完全排除。

③匹配题：测题结构上包含两个部分，一个为一组刺激项目，另一个为一组反应项目，通常是在后者中选出与前者相适合的项目。

A.优点：覆盖面较广、编拟起来也相对容易。

B.缺点：对选项的同质性要求较高，无法测量出被试的言语表达能力和概括、组织能力。

(2) 主观题

包括填空题、简答题、应用题、论文题、联想题、操作题等。这里着重介绍各类题型中最易于编制的简答题。

简答题要求被试用一个正确的短语、句子或较为简单的一段文字来完成测题。

①优点：编写灵活；不易受猜测的影响；可测量各种层次的知识目标和能力，应用较为广泛；不用考虑选项之间的同质性问题。

②缺点：不能测量复杂的知识和能力，评分不够客观且费时。

(四) 测验标准化

测验标准化是指测验的编制、施测、评分以及解释测验分数的程序的一致性。

1.测验内容标准化

首先要保证所有受测者施测相同或等值的题目。

2.测验编制标准化

在确定测验维度、题目编制、测验信度和效度分析、项目分析等方面严格按照规定程序进行。

3.施测过程标准化

(1) 具有相同的测验情景（采光、房间布置等）；

(2) 相同的指导语（说明测验目的及如何做反应）；

(3) 相同的测验时限（不同测验对时间限制要求不同）。

4.测验评分标准化

一般要求评分者之间的一致性达到90%以上，便可认为是客观的。具体包括：

(1) 对反应及时清楚地记录，避免遗忘；

(2) 有计分键（即标准答题卡）；

(3) 将受测者反应和计分键比较，确定其反应应得分数。

5.解释分数标准化

某一测验分数只有与一定的参照标准相比较，才能显现其所代表的意义。

(五) 测验等值

1.含义

通过对考核同一心理品质的多个测验形式做出测量分数系统转换，使得这些不同测验形式的测验分数之间有可比性。

2.测验等值的条件

(1) 同质性：被等值的不同测验形式所测的必须是同一种心理品质且内容与范围基本相同；

(2) 等信度：被等值的不同测验形式必须有相等的测验信度；

(3) 公平性：考生参加被等值的不同测验形式中的任何一个测试，等值后的结果都是一样的；

(4) 可递推性：由测验 X 与测验 Y 之间的等值关系，以及测验 Y 与测验 Z 之间的等值关系，可以递推出测验 X 与测验 Z 之间的等值关系；

(5) 对称性：从等值的两个测验中的任何一个出发，得到的等值结果都应该是相等的；

(6) 样本不变性：两个测验的转换关系是内在的，不随所使用样本的变化而变化，也有学者将前四个条件合称为公平性。

3. 计算方法

(1) 等百分位等值

如果两个分数对于任何一个被试群体都有相同的百分等级，那么这两个分数就是等值的。具体操作是寻找与 x 分数有相等百分等级的 y 分数可使用作图法和计算法。

(2) 线性等值

如果两个分数在各自群体内的标准分数相等，则这两个分数是等值的。

$$\frac{x - \bar{x}}{S_x} = \frac{y - \bar{y}}{S_y}$$

$y = Ax + B$ ，AB 为等值常数， $A = \frac{S_y}{S_x}$ ， $B = \bar{y} - A\bar{x}$

4. 等值结果的表示方法

(1) 表列法；

(2) 公式法；

(3) 图示法。

5. 基本概念

(1) 经典理论等值与项目反应理论等值

① 经典测量理论等值以经典测量理论为指导；

② 项目反应理论等值以项目反应理论为指导，对等值的要求更宽松，等值结果更准确。

(2) 测验分数等值与项目参数等值

① 前者指各测验间原始分数的等值；

② 后者指各测验间项目参数的等值，且只能在项目反应理论指导下才行。

(3) 水平等值与垂直等值 $\bar{y} = A\bar{x}$

① 前者指被等值的测验难度水平、受测团体的能力都相似；

② 后者指被等值的测验难度水平、受测团体的能力水平都不同。

(4) 等值标准误差与等值偏差

① 前者指抽样引起的等值误差，等值分数越趋于分布两端，其值越大；

② 后者指等值处理方法不当引起的等值误差。

(5) 锚测验与数据平滑法

① 锚测验：测验等值设计中，用来关联两个待等值的测验，以寻找两种形式的等值关系的一组测验；其制定要求包括：与原测验一样测量相同的心理品质、等信度、长度，不小于原测验的 1/5。

②数据平滑法：用来使分布曲线趋近平滑的统计技术，以减少误差对等值的影响。

6.步骤

- (1) 定等值目标；
- (2) 作等值设计；
- (3) 采集数据；
- (4) 选操作定义；
- (5) 算等值结果；
- (6) 评价。

二、心理测验的施测

(一) 测验的设计

测验的设计是指对测验的结构和内容、进行测验可达到的目的、测验结果可提供的信息、测验的适用范围、测验的心理测量学指标等进行总体设计。进行心理测验前，主试应充分了解测验的设计，根据考查的目标、受测者的特点、测验设计等来选择合适的测验。测验的使用一般应按照测验手册严格执行。

总体上，测验过程包括选择测验、测验实施、测验评分、结果报告四个过程。

(二) 影响因素

1.施测前：准备好测验材料，熟练掌握施测手续，熟记测验指导语并能用口语清楚流利地说出来。

2.指导语：其主要作用是使受试者按正确的形式对题目做出反应，主试和指导语都应保持中立，不倾向于答案的任何一方。

3.测验情景：测验场地（通风、光线、噪音），座位，答案纸型等都会影响测验分数。

4.测验焦虑：会影响结果的真实性。测试时要稳定被试的情绪。能力越高，测验焦虑越低；越渴望得到高分的人，测验焦虑越高。竞争性测验焦虑高，经常接受测验的人的焦虑低。轻微焦虑会增加测验效果，焦虑太高或毫无焦虑则会降低测验效果。

5.与受测者建立良好的协调关系：保证他能按指导语行事。

6.评分技术：要标准严格，采用计分键。

(三) 测验分数的解释

1.如何看待分数

测验结果须针对具体测验、受测者情况以及测验的具体情况。解释分数时要注意：

- (1) 测验的具体功能；
- (2) 参考其他有关资料，防止片面极端；
- (3) 将测验分数当作是“一段分数”，而不是一个点；
- (4) 明白测验常模和效度的局限；
- (5) 明白不同测验分数间不能直接比较。

2.如何报告分数

- (1) 语言能被理解且保证当事人知道测验目的；
- (2) 告诉当事人分数并不是一个精确的值，不是决策的标准；
- (3) 对待低分解释时要小心；

(4) 测验结果保密, 并提供相应指导。

三、测验常模

(一) 分数转换

1. 原始分数与导出分数

(1) 原始分数: 测验中直接获得的分数。

(2) 导出分数: 在原始分数转换的基础上, 按照一定的规则, 经过统计处理后获得的具有一定单位和参照点且 $Z = \frac{X - \bar{X}}{s}$ 可以相互比较的分数量表或符号系统。

2. 百分等级

(1) 含义

指一个群体的测验分数中, 得分低于这个分数的人数的百分比(计算见心理统计学部分)。百分等级分数是应用最广的导出分数。

(2) 评价

①优点: A. 具有可比性; B. 易于计算; C. 解释方便; D. 较适用于各种被试和各种性质不同的测验; E. 百分等级不受原始分数分布状态的影响。

②缺点: A. 不具有可加性; B. 单位不等, 尤其在分布的两个极端; C. 百分等级只具有顺序性, 而无法用它来说明不同被试之间分数差异的数量; D. 百分等级是相对于特定的被试团体而言的, 解释时不能离开特定的参照团体。

3. 标准分数及变式

(1) 标准分数

具有稳定性、明确性、可加性、可比性等优点, 但存在计算繁杂、有负值和零、有小数、在进行比较时须满足数据的原始形态相同等缺点(详见心理统计学部分)。另外, 当两个原始分布形态不相同, 可进行非线性转换, 将原始分数都转换成正态分布, 具体步骤分为: ①原始分数转为百分等级; ②百分等级查表得出 Z 值。

(2) 标准分数的变式

小数和负数的存在使线性的和正态化的标准分数在计算和解释上有些不适, 为此, 要将标准分数作线性变换。一般的转化形式为: $y = m + k(Z)$ 。其中, y 为转化后的分数, 而 m 和 k 为常数, m 为转化后新的分数分布的均值, 而 k 则为其标准差, 具体包括:

①T 分数

$T = 50 + 10Z$, 均值为 50, 标准差为 10。

②CEEB 分数(美国大学入学考试委员会使用的标准分数)

$CEEB = 100Z + 500$, 均值为 500, 标准差为 100。

③离差智商

$IQ = 15Z + 100$, 均值为 100, 标准差为 15。

④EPT(我国一种出国人员英语水平考试的标准分数)

$EPT = 20Z + 90$, 均值为 90, 标准差为 20。

各种变式均具有相等单位; 在正态分布下, 可以利用正态分布表将各种导出分数与百分

等级分数作换算；正态分布下，运用某种变式分数可以将几个测验上的分数作直接的比较。即使是非正态分布，也可运用由正态化的 Z 分数转换而得到的变式分数进行直接比较分析。

但是，变式分数的解释过于抽象；在非正态分布下，不同分布形态的变式分数不能相互比较和求和。

4. 标准九分

标准九分数是将原始分数分成九个部分的标准分数系统。若原始分数服从正态分布，它是以 0.5 个标准差为单位，将正态曲线下的横轴分为九段，最高一端为 9 分，最低一端为 1 分，中间一段为 5 分。除两端外，每段均有半个标准差宽。

(二) 分数合成

1. 含义

分数合成是指将几个分数或几个预测源组合起来，以获得一个合成分数或做出总的预测；具体包括项目的组合、分测验或量表的组合以及测验或预测源的组合。

2. 合成方法

(1) 临床诊断法：根据经验而依靠直觉合成分数。具有高度综合性、灵活性，但不够客观、数量指标不精确、要求判断者受过训练并具有丰富经验。

(2) 加权求和法：若各测验所测特质间有互相代偿作用，这些测验的分数又是连续型资料，并能大体同时获得时，可以采取加权求和的方法。具体加权方式包括：单位加权（测验分数直接相加）、等量加权（转换为 Z 分数后相加）、差异加权（转换为 Z 分数后，根据不同测验的权重进行相加）。

(3) 多重回归法：研究一种现象和其他多种现象在数量上相互联系和相互制约的统计方法。前提是所测特质间具有某种程度的互偿性，预测源与效标间有线性相关且都是连续型数据，可以同时取得两种数据。

(4) 多重划分法：在各个特质上都确定一个标准，从而把成绩划分为合格和不合格两类。只有通过了前一个测验，才能继续实施后一个测验；只有每个测验都合格，总要求才算合格，又称连续栅栏法。这种方法通常将最有效的预测源或测验放在前面。

当所测特质间不具有互偿性，或者预测源和校标间无线性相关时才使用。

(5) 其他：完形计分、轮廓分析等。

(三) 常模及常模团体

常模参照分数是将受测者的成绩与具有某种特征的人所组成的有关团体作比较，根据一个人在该团体内的相对位置来报告他的成绩。这里，用来做比较的参考团体叫常模团体，常模团体的分数分布叫作常模。

1. 常模

根据标准化样本（常模团体）的测验分数，经过统计处理而建立起来的具有参照点和单位的测验量表；是关于原始分数、导出分数和常模团体的有关具体描述。

2. 常模团体

(1) 含义

是具有某种共同特征的人组成的一个群体。若群体较大，常模团体则是该群体的代表性取样，因此又称作标准化样本。

(2) 注意事项

- ①群体构成的界限必须明确；
- ②样本必须具有代表性，要注意克服取样误差，遵循随机化原则，用统计方法抽取样本；
- ③取样过程必须详尽描述；
- ④样本大小要适当，一般来说，样本最好有 30 至 100 人，如果是全国性常模，一般应有 2000 至 3000 人；
- ⑤要注意常模的时间性；
- ⑥一般常模与特殊常模结合。

(3) 呈现常模的方法

- ①转化表：主要包括简单转换表和复杂转换表两种。如韦克斯勒智力量表。
- ②剖析图：又称剖面图，可将一套测验中不同维度或分测验分数用图表(形)表示出来，可从剖析图上直观地观察被试在各个分测验上的表现及相对位置。但要求各个分测验采用同一个常模团体，否则无法比较。如卡特尔 16 种人格因素问卷 (16PF)。

3.常模的编制

- (1) 确定群体，抽取常模团体；
- (2) 对常模团体进行施测，获得测验分数及分数分布；
- (3) 制作常模分数转换表，即常模量表。

4.常用的常模参照分数

常模一般可分为两类，一类是发展常模量表，是指某类个体正常发展过程中各个阶段的一般水平，是将被试得分与不同发展水平的人作纵向比较；另一类是组内量表，是指某类个体在测验所测特征上的一般表现水平，是将被试得分在一个团体内作横向比较，主要有百分等级常模与标准分数常模两种。

(1) 发展量表

- ①年龄常模：各年龄儿童的正常水平。

通过将被试的成绩与相应年龄组学生的正常水平作比较，判断其水平。

A.要素：区分不同年龄组的题目、常模团体、常模表（得多少分该归入哪个年龄的对照表）。

B.评价：容易理解，但是非等单位，如 3 岁和 4 岁间的差异不等于 10 岁和 11 岁间的差异。

- ②年级常模：不同年级学生在某种测验上的正常水平。

通过将被试的成绩与相应年级学生的正常水平作比较，判断其水平。

(2) 商数

- ①教育商数

$$EQ = \frac{\text{教育年龄}}{\text{实际年龄}} \times 100 = \frac{EA}{CA} \times 100$$

- ②智力商数（比率智商）

$$IQ = \frac{\text{智力年龄}}{\text{实际年龄}} \times 100$$

③成就商数

$$AQ = \frac{\text{教育年龄}}{\text{智力年龄}} \times 100 = \frac{\frac{\text{教龄}}{\text{实际年龄}}}{\frac{\text{智龄}}{\text{实际年龄}}} \times 100 = \frac{EQ}{IQ} \times 100$$

(3) 百分等级常模

基于某个常模团体，为测验的原始分数与百分等级之间建立起对应关系的组内常模。百分等级指出的是个体在常模团体中的相对位置。

这种常模由原始分数、百分等级、常模团体组成，常以转化表形式出现，包括简单转换表和复杂转换表。使用者可将原始分数转化为相应的百分等级，也可根据百分等级找到原始分数。

(4) 标准分数常模

基于某个常模团体，将测验的原始分数转换成标准分数 Z 或导出分数 T，从而反映每个分数在常模团体中相对位置的一种组内常模。

(5) 剖面图

将一套测验中几个分测验的分数用图形表示出来。可直观看到被试在各个分测验中的表现及相对应的位置。

四、标准参照测验

(一) 概述

1. 含义

又称目标参照测验（戈莱塞，1971 年提出），即根据某明确界定的内容范围而镇密编制的测验，并且被试在测验所得结果也是根据某一明确界定的行为标准直接进行解释的，比如：四六级考试，要求达到 425 分才算通过。（这是郑日昌老师 1995 年的观点。大家现在参加的四六级考试是常模参照测验。）

2. 作用

了解个体在所规定的测量内容上的行为水平，其出发点是个体本身的绝对水平，而不再是个体间的水平差异（常模参照测验关注）。

(二) 信效度

1. 信度

该测验并不采用常模参照测验中运用相关系数计算出来的信度系数，原因是相关系数易受分数分布的影响，而标准参照测验关注的是个体本身的绝对水平，不考虑分数分布。

(1) 分类一致性信度：考察同一被试团体在同一测验的两次施测中是否被分在同一类中（如两次都及格或两次都不及格）， $P_0 = (a+d) / N$ 。a 为两次都及格的人数，d 为两次都不及格的人数； $N = a+b+c+d$ ，为总人数。

(2) 荷伊特信度：采用方差分析法，找出个体真实水平的变异在总变异中占的比例。此方法不需考虑分界点问题，应用更普遍。

2.效度

(1) 内容效度

评估任一测验的内容效度，都依赖两个条件：

- ①测验有明确界定的内容范围；
- ②对测验每一题目的内容效度的分析。

标准参照测验一般来说有相对比较确定的内容范围，可用命题细目表表示。同时，也可以采用专家评定的方法对题目效度进行分析，从而保留有效题目，删除无效题目。其方法类似常模参照测验。

(2) 效标关联效度（决策效度）

标准参照测验的效标关联效度分析方法与常模参照测验的方法没有太大差异，不同之处在于统计指标上。由于标准参照测验不适合使用相关系数作为指标，因此提出“决策效度”的概念来评估目标参照测验的效标关联效度。决策效度就是在预测原测验和效标测验中击中 and 正确否定的比例之和（见表）。决策效度= (a+d) / d

		原测验	
		掌握	未掌握
效标测验	掌握	a	b
	未掌握	c	d

(3) 结构效度

关于标准参照测验的结构效度，目前尚未得到较大关注。

(三) 项目分析

1.测验的预测

在获得预测资料之后才能对项目的难度和区分度进行量化分析。

- (1) 前测——后测方法：假设前测为“未掌握组”，后测为“掌握组”。
- (2) 已接受教学组——未接受教学组方法：假设未接受教学组为“未掌握组”，已接受教学组为“掌握组”。
- (3) 对照组方法：由教师评定是否掌握知识。

2.难度分析

与常模参照测验一样，常以通过率（P）表示。有人认为没有必要在标准参照测验中分析测验的难度。

3.区分度分析

(1) 难度差值

①掌握组——未掌握组鉴别指数（D）： $D=P_A - P_B$ ； P_A 和 P_B 表示掌握组和未掌握组的通过率， $-1 \leq D \leq 1$ 。

②个人获得指数（ D_{IG} ）：采取前测——后测法时，可以获得在前测中错误回答而在后测中能够正确回答的被试人数的比例。其值在 0~1 之间。

(2) 相关系数

项目得分和测验总分之间的一致性程度，常被用作项目区分度的指标，其值在-1~1 之间。

4.内容范围的确定

内容包括欲测量特质中蕴含的全部行为。

(1) 内容范围具有明确边界；

(2) 每一内容范围均可分类，每类可再分为更细小的类；

(3) 界定结果常以命题双向细目表（包括教学目标、教学内容以及每一内容和每一目标结合后所占的比重）呈现。

5.内容效度分析

常用专家评定法，如填写内容评定表、五级量表。

（四）分数解释

对标准参照测验的分数进行解释时，要特别注意分界点的确定。分界点的确定方法有利弊，大体包括：

（1）专家判定法

在测验内容范围明确界定的基础之上，由专家来判断处于临界水平的被试在每一题目上正确回答的可能性，进步以此为标准以确定分数分界点。

所谓临界水平的被试，是指那些刚由未掌握水平转入掌握水平的被试，这些被试实际上是在专家的想象中虚拟出来的。具体的评定方法有以下两种：

①Nedlsky 法

由 Nedlsky 提出，针对由多重选择题组成的测验而言，由专家来判断处于临界水平的被试在每一题目上有能力排除的错误选择项，从而计算其正确回答的可能性，再求出每一题目上正确回答的可能性之和，即为测验分数分界点。

若请若干专家同时评定，则可以采用这些专家所评定及格线的平均值作为最终及格线。

②Angoff 法

由 Angoff 提出，由专家直接判断处于临界水平的被试在某测验的每一题目上正确作答的可能性（记为 P_i ），设每题的满分为 F_i ，则该测验的分数分界点（记为 λ ）为 $\lambda = \sum P_i F_i$ 。

若有多位专家同时评定，则以这些专家评定的平均及格线作为测验最终及格线。

比较 Nedlsky 法和 Angoff 法，前者使专家的评定受到限制，而 Angoff 法中的 P_i 则可在 0~1 之间任意取值，而且适用于各种题型。

③Bookmark 法

基于项目反应理论，以 Angoff 法为基础的专家评定法。

（2）效标组预测法

①临界组法。由专家判定和选择一组正处于临界水平的被试，将测验施测于该组被试，计算他们在测验上的平均成绩，以体现测验内容范围所要求的临界水平，因而可以视之为测验分数分界点的估计值。

采用此法的困难在于：A.必须先随机选取大量被试作为候选；B.被试是否处于临界水平很难找到客观而统一的标准，非常抽象、主观。

②对照组法。同样要先由专家来选择被试，确定两组被试，一组被明确判定为掌握组，另一组则被明确判定为非掌握组，那些不太容易被判定的被试一概剔除。两组被试原始分数分布曲线的交叉点即为测验分数的分界点。

更合理的做法：是选取若干对对照组，取每对对照组交叉点分数的平均值作为测验分数的分界点。

（五）常模参照测验与标准参照测验

1.相同点

- （1）均根据定的法则对人的行为和心理进行量化；
- （2）两种测验都具有间接性、相对性及客观性等特点；
- （3）均有三个要素，即行为样组、标准化和客观性；
- （4）二者在信度、效度、难度和区分度的指标中，有很多是相同的（如荷伊特信度、内容效度、效标效度、通过率、鉴别指数法、相关法等）；
- （5）教育测验应该注重两种目标并举。

2.不同点

（1）目的不同

常模参照测验将被试与常模团体相比较，旨在评价被试在团体中的相对地位；标准参照测验则是将被试与某一绝对标准比较，旨在评价被试有无达到该标准。

（2）信效度方面在常模参照测验中常用相关系数作为信度指标，以测验与效标之间的相关系数作为效标关联效度的指标；标准参照测验中则采用荷伊特信度和分类一致性信度，以“决策效度”作为效度指标。

（3）难度、区分度方面

常模参照测验要求定的难度，区分度关心对心理品质的区分；标准参照测验不太重视难度，区分度关心对在其内容范围上的已掌握者和未掌握者做出最大限度的区分。