

教育测量与评价

第一章 教育测量与评价的学科发展

第一节 教育测量与评价的基本问题

一、教育测量与评价的含义

1. **测量**：从广义上是根据某些法则与程序，用数字对事物在量上的规定性予以确定和描述的过程。

2. 测量的**要素**：一是必须依据某些科学原理与法则，发展出合适的量具或制定出科学的测量方案；二是必须有意义相对明确的测量单位；三是用数字对事物在量上的规定性予以确定，就需要有一个测量或计算的起点，即参照点。参照点不同，其策略结果也就不同，且测量结果之间也无法进行直接比较。

参照点分绝对零点和相对零点。测量三个基本要素：量具、单位、参照点。

3. 教育测量的**含义**：针对学校教育影响下学生各方面的发展，侧重从量的规定性上予以确定和描述的过程。教育测量反馈的是关于课堂教与学两方面的信息。

4. 教育测量的**特点**：总的来看，教育测量是属于精神特性的测量。

与物理量的测量相比，其具有（1）：间接性和推断性。目前只能通过人的外显行为或通过人对来自外界的一些刺激所作出的反应，对人的知识技能、智力水平、思维品质、创造能力、心理素质、情感态度、思想道德等作出间接性、推断性的测量。（2）测量对象的模糊性和测量误差的不可避免性（3）量表具有多样性，结果具有相对抽象性。量表根据测量的精确程度，从低级到高级可分为称名量表、顺序量表、等距量表和比率量表四种水平（各自定义）。测验的原始分数经过统计处理后推导出一种新的量表分数，此时分数“零”有时也是一个相对的零点。但绝大多数教育测量的内容不可穷尽，故大多数教育测量往往没有绝对的零点。

二、教育评价的基本问题

1. 评价的**含义**：广义的评价泛指衡量、判断人物或事物的价值。

2. 教育评价的**概念**：格兰郎德（N. E. Gronlund）认为，评价是为了确定学生达到教学目标的程度，收集、分析和解释信息的（课堂）系统过程；评价包括对学生的定量描述（测量）和定性描述（非测量）两方面。根据格兰朗德的观点，评价总是包括对测量结果需求程度的价值判断。一个完整的评价计划可用公式表达： $评价 = 测量（定量描述） + 非测量（定性描述） + 价值判断$ 。

广义的教育评价是指按照一定的价值标准和教育目标，利用测量和非测量的

种种方法系统地收集资料信息，对学生的发展变化及其影响学生发展变化的各种要素进行价值分析和价值判断，并为教育决策提供依据的过程。

三、教育评价相关概念辨析

1. 教育测量与教育评价

教育测量是针对教育效果或者针对学生各方面的发展予以测量和描述的过程，旨在获得有一定说服力的数量事实，是一种以量化为主要特征的事实判断。教育评价是根据一定的标准，对教育事物或现象的价值进行系统的调查，在获取足够多的资料事实（定性资料与定量资料）基础上，作出价值分析和判断。

2. 教育评价与评估

3. 教育测验与考试

第二节 教育测量与评价的发展历史

一、中国是考试制度的发源地

二、中国科举制度的世界地位

1. 科举制度始于隋炀帝大业二年（606），直到清光绪二十九年（1903）举行最后一科进士考试后宣布结束。

三、教育测量学科的诞生

1864年，英国教师费舍（G. Fisher）编制《作业量表集》，作为教师以后评定学生各科成绩的参照标准。美国心理学家桑代克于1904年出版论著《精神与社会测量导论》，首次较为系统地介绍了教育统计方法及编制测验的基本原理。

四、教育测量运动的开展

20世纪初，在桑代克的影响下，美国的教育测量运动轰轰烈烈地开展。英国心理学家斯皮尔曼发表了“能力结构二因素学说”。伍德沃斯的第一个标准化人格测验极大推动了心理测验在美国的发展和应用，对美国当时的教育测验发展起了重要的促进作用。

五、美国的“八年研究”

1. “八年研究”**基本原则**：第一，应当帮助学生学习的什么，即教育计划中，应当帮助学生发展哪些思维、感觉和行动的方式。第二，为帮助学生达到教育目标，应当提供什么学习经验，即怎样帮助学生学习的预期的东西。第三，怎样组织学习经验，以使这些经验的累计效果最大，即怎样的学习顺序和学习经验的整体计划有助于学生内化其所学之物，并将其应用于适当的环境。第四，怎样评价方案效果，即随后采取什么步骤，以不断检查期望的学习活动所发生的变化。

2. 基于“八年研究”，教育评价思想与理论的发展。泰勒在“八年研究”中提出了评价活动原理和课程编制原理和课程编制原理。基于泰勒的理论建树，形成了泰勒的“行为目标评价模式”，泰勒因此被称为当代教育评价之父，这个时期的教育评价在美国教育史上称为“泰勒时期”。

第三节 教育测量与评价的学科地位和作用

一、教育测量与评价是现代教育科学研究的三大领域之一

1. 美国教育测量与评价理论研究相比教育科学体系中其他一些研究领域的主要表现：第一，创办了大量有关教育测量与评价方面的研究杂志，如《教育测量研究》等；第二，出版了大量有关教育测量与评价方面的学术著作与教科书，如《教育测量》等；第三，许多学术团体都加盟了对教育评价的研究，成立了一些全国性的教育评价协会，如全国教育测量委员会等。

二、教育测量与评价在教育改革中具有重要作用

1. 教育改革常以教育测量与评价的改革作为突破口
2. 考试与评价改革成为基础教育新课程改革的重要内容

第二章 教育测量与评价的类型和功能

第一节 教育测量与评价的主要类型

一、按测量与评价在教学中运用的时机分类

1. 形成性评价。形成性测量与评价是在教学过程中经常实施的，在性质上大致相当于现在的中小学单元检测。
2. 诊断性测量与评价
3. 终结性测量与评价

二、按解释测量结果或评价结果时的参照点分类

1. 常模参照测量与评价
2. 标准参照测量与评价
3. 潜力参照测量与评价

三、按测量与评价被试行为表现的性质分类

1. 最佳行为测量与评价
2. 典型行为测量与评价

四、按测量与评价的内容分类

1. 智力测量与评价

2. 能力倾向测量与评价
3. 成就测量与评价
4. 人格测量与评价

第二节 教育测量与评价的主要功能

一、实现教育判断的功能

1. 测量评定的功能
2. 事实判断的功能
3. 价值判断的功能
4. 问题诊断的功能
5. 区分选拔的功能

二、改进教师教学的功能

1. 了解学生的起点行为
2. 作为改进教学的参考
3. 作为补救教与学的依据
4. 确保教学目标的达到

三、促进学生学习的功能

1. 明确目标，增加学习动力
2. 了解自我，调整学习计划

四、行使教育的功能

1. 对教师的管理
2. 对学生的管理
3. 对教育目标和质量的管理
4. 对教育过程的管理
5. 对学校的管理

第三章 教育测量与评价的质量特性

第一节 教育测量与评价的信度

一、信度的意义

信度：指的是测量结果的稳定性程度，记为 r_{xx} ，即如果用同一测量工具反复测量同一种特质对象，则多次测量结果间的一致性程度。

二、信度的统计定义

在测验实施过程中直接得到的分数，称为观察分数，记为 X ；把被试在所测特质上客观具有的水平值，称为真分数，记为 T ；观察分数与真分数的差就是测量误差，记为 E 。三者的关系可表示为： $X=T+E$ 。

信度的实际意义：信度是一个被试团体在测验 X （A 卷）上的观察分数与在测验 X 的任意一个“平行测验” X' （B 卷）上的观察分数的相关系数。即：

$$r_{xx}=\rho_{xx'}$$

三、信度的估计方法

- 1.重测信度。适用于异质性测验。
- 2.复本信度。也成为平行测验。指的是两个平行测验测量同一批被试所得结果的一致性程度，其大小等于同一批被试在两个复本测验上所得分数的相关系数。实施复本信度的两种方式：一是在同一时间连续施测，可反映出测验内容造成的误差的多少，称为等值性系数；二是间隔一段时间后测验，不仅反映出测验内容的抽样误差，而且可反映被试本身状况的改变，同时兼顾试题抽样与时间影响的信度，称为等值稳定性系数。
- 3.同质性信度。包括分半信度、库德-理查逊信度（也称库-理信度）和克隆巴赫 α 系数。

第二节 教育测量与评价的效度

一、效度的意义

效度：效度是指一个测验或量表实际能测出其所要测量的特性的程度。

二、效度的统计定义

被试的观察分数是由真分数与误差分数构成，即 $X=T+E$ 。实质上，真分数还可进一步分解为两部分：目标真分数（记为 V ）和非目标真分数（记为 I ）。目标真分数指的是反映被试某种心理特质真正水平的数值。非目标真分数，指被试在某种心理特质测量量表上表现的与测量目标无关的稳定测值。故观察分数可表示为： $X=V+I+E$ 。测量效度实质上指一次测量出目标真分数的程度。

三、效度分类

1. 内容效度
2. 结构效度
3. 效标关联效度

第三节 教育测量与评价中题目（项目）的难度

一、难度定义

教育测量与评价中题目或项目的难度,值被试完成题目或项目任务时所遇到的困难程度。定量刻画被试作答一个题目所遇到的困难程度的量数,称为题目的难度系数,也常称为难度值,用符号 P 表示。

第四节 教育测量与评价中题目(项目)的区分度

一、区分度的意义

题目区分度是题目区别被试水平能力的量度,常记为 D 。

二、区分度的计算

1. 相关法
2. 高低分组法

第五节 教育测量与评价方案的可行性

- 一、科学性、公平性和可行性(展开阐述)
- 二、针对性、区分性和简洁性(展开阐述)

第四章 编制教育测验的一般原理与方法

第一节 测验题目类型与测量功能

测验题目基本上可分为两大类:客观性试题和主观性试题。客观性试题主要包括选择题、填空题、简答题、是非判断题等,它们的答案可在测验前就已准备好,不同评分者各自独立评分,所得结果基本上相同。主观性试题主要包括论述题、操作题和作文题等。它们适合于测量较高层次的教学目标,尤其是适合于测量综合、评价等目标层次。

二、主观性试题的类型及其编写要领

一、论述题

1. 定义:是向被试提出问题,需要被试用自己的语言组成一份较长答案的试题。其最大的特点是被试在回答问题时,有较大的自由度,可以充分地运用所学的知识,并且可以加上自己独特的见解。
2. 编制原则:
 1. 试题应该用来测量较高层次的教学目标,像综合、评价等目标层次,限于那些客观题不能测量的学习成就。
 2. 要明确而系统地陈述问题,使被试能清楚地了解题目的要求。
 3. 一般采用答案具有统一定论的试题。
 4. 当有多个论述题时,可设置一个必答的论述题,剩下的题目可允许被试选择作答。
 5. 为避免被试将时间集中在某一个其不会做的题目上而影响对其他题目的回答,并因此而影响考试成绩的真实性,最好在题目中能给出回答本题所需的参考时间。

二、作文题

1. 定义：实质上一种论述题，是对人的逻辑思维、形象思维、书面表达等多种能力的一种综合考查。

2. 编制原则：（1）根据考试目的、考试对象来确定作文文体和要求。（2）根据社会需要、现实生活和学生实际设计命题，作文命题要直接测量语文素养和书面表达能力。（3）作文命题不能过于抽象，要让学生有话可说，有内容可写，有思路可走，有发挥的余地。（4）避免材料或话题出现测验偏倚，在确定选材和写作意图时，要考虑对所有被试都是公平的，并且试题应符合被试的心理特质。

三、表现性测验题

1. 定义：指以行动、作品、表演、展示、口头回答、操作、写作、科学研究、制作等更真实的行为表现，来测量与评价学生在真实情景或模拟情景下的所知与所能的一种测验类型。

2. 根据真实程度分，表现性测验类型可分为：（1）纸笔的表现性测验。（2）辨认的表现性测验。（3）结构化的表现性测验。（4）模拟情景的表现性测验。（5）工作样本的表现性测验。

3. 编制原则：（1）测验材料和要求应与日常教学内容和目标有关联，关注重要的学习结果。（2）将考察重点放在测量复杂和高层次能力上。（3）将干扰测验目的的无关困难降到最低。（4）测验尽可能选择真实性的情景或问题。（5）提供必要的操作指导语，让学生了解作业任务和期望内涵。（6）指导语简单清晰，向被试明确说明实际的情境、任务以及评价的标准。（7）确定科学合理的计分方法。对操作行为的计分通常要将程序、作品两者结合起来。

第二节 测验目标与教育目标研究

二、教育目标分类研究

教育目标可分为终极教育目标、中程教育目标和具体教育目标。为满足教育测验的需要，布卢姆等人首先把教育目标分为认知、情感和动作技能三个领域。

1. 认知领域。布卢姆将认知领域目标分为六个类别，按照由低级到高级的难易程度形成一种递进的等级关系，前一类别是后一类别的基础，后一类别又涵盖了前面的类别。

（1）知识 最低等级的认知目标。此处的知识是指对具体事物和普遍原理的回忆，对方法和过程的回忆，或者对一种模式、结构或框架的回忆，知识目标强调记忆的心理过程。

(2) 领会 最低层次的理解，指个人把某种材料与其他材料联系起来，也不必弄清它的最充分的含义，便知道正在交流什么，并能够运用正在交流的这种材料和观点。

(3) 运用 指将抽象的概念用于特定的和具体的情境。这些抽象的概念，可能是以一般的观念、程序的规则或概括化的方法等形式表现出来的，也可能是那些必须记住的和能够专门运用的原理、观念和理论。

(4) 分析 指把材料分解成各个组成部分，弄清楚各部分之间的相互关系及其构成的方式，以指出那些用来传递意义或确定交流结果的技术和手段。

(5) 综合 指将各种要素和组成部分结合起来，以形成一个整体。它是对各种要素和组成部分等进行加工的过程，是一个用这种方式将它们组合起来，以构成一种原先不那么清楚的模式或结构的过程。

(6) 评价 指为了某种目的，对观念、作品、答案、方法和资料的价值及符合准则的程度作定量和定性的判断。

2. 情感领域。布卢姆与克拉斯沃尔等提出了情感领域的教育目标分类。将情感领域的目标分为接受、反应、价值评价、组织和由价值或价值复合体形成的性格化五个类别。

(1) 接受 指学生愿意注意某些现象或刺激。这一目标包括觉察到某一事物的存在，愿意接受某种特定刺激和有选择的注意这个事物的三个亚类。

(2) 反应 指学生受到动机的充分驱动，积极地注意某种观念，并伴随着主动参与行为。这一水平上，学生不仅注意到了某一种现象，还以某种方式对它作出了反应。

(3) 价值评价 指接纳并赋予某一客体、现象或行为价值。

(4) 组织 是将各种不同的价值结合起来，解决它们之间的冲突并开始建立一个内在的、一致的价值体系，其重点是价值的比较、关联和综合。包括价值的概念化、价值体系的组织化。

(5) 由价值或价值复合体形成的性格化 这是情感领域教育目标的最高境界。在这一层次上，各种价值在个体价值结构中的位置已经确定，已形成了一个价值体系，能在相当长的时间内控制个体的行为，使其在相当长的时间内能形成内化的价值行为。

3. 动作技能领域 辛普森将动作技能领域目标分为：知觉、准备、指导反应、机械化、复杂反应、适应性、独创性。

三、确定测验目标应注意的问题

1. 深入分析课程目标
2. 正确选择目标领域和类别
3. 正确理解各类目标之间的关系

第三节 测验蓝图设计与测验编制要领

一、设计测验的基本考虑

1. 确定测验的目的
2. 确定测验的属性
3. 明确测验的性质和用途
4. 明确测验的对象
5. 分析测验的目标

二、设计测验蓝图

1. 作用：第一，保证测验试题是所测量的教学内容的代表性样本，且能反映出各个部分内容之间的相对重要性，以便测验内容取样适当，提高测验的效度。第二，规划代表不同的知识内容和学习水平的各类测验目标的比例分配，在保证测验效度的同时保证测验的难度合理。
2. 步骤 （1）确定测验内容要目 （2）确定该科目应考察的目标层次 （3）确定各项测验内容的要目下的权重 （4）形成命题双向细目表

三、测验的编制和组织

1. 选定测验的材料
2. 编写测验试题（命题）
3. 试测与题目分析
4. 合成测验（组卷）
5. 编制复本
6. 编写测验手册

第五章 制定教育评价表的一般方法与步骤

第一节 教育评价表的基本类型与制定原则

一、教育评价表的基本类型

1. 概括性问题教育评价表 是根据评价目标概括提出一系列问题加以系统了解的评价表式。

2. 指标结构性教育评价表 是根据评价目标逐层分解评价指标，形成具有评价指标体系、指标权重结构以及定性定量具体评价标准的教育评价表。

二、制定教育评价表的基本原则

1. 教育规律性和导向性原则
2. 科学性和可操作性原则
3. 超前性和发展性原则

第二节 教育评价表的构成及编制方法

一、教育评价表的构成要素

1. 评价指标
2. 指标权重
3. 评价标准

二、制定教育评价表的方法

1. 确定教育评价的对象和目标 即要确定评价对象，即确定评价谁（什么）的问题。评价对象可以是人（如教师、学生等），也可以是物（如学校评价、教材评价等）
2. 初拟评价指标 具体方法：（1）头脑风暴法（2）因素分解法（3）理论推演法
3. 筛选评价指标 方法：（1）经验法（2）调查统计法
4. 确定评价指标权重 方法：（1）关键特征调查法（2）两两比较法（3）专家评判平均法（4）倍数比较法
5. 设计教育评价标准 方法：（1）分解教育评价表中指标所包含的主要内容（2）确定标度（3）确定等级数量
6. 整合、修改与完善教育评价表

第六章 教育测验的常模及其建立方法

第一节 教育测验常模的意义与类型

一、测验常模的意义

在教育测量学中，测验的常模，简称常模，是指一个代表性的样组在某种测验上的表现情况，或者说，是一个与被试同类的团体在系统测验上得分的分布状况与结构模式。

二、测验常模的主要类型

1. 年龄常模

2. 年级常模

第二节 百分等级常模及建立方法

一、百分等级常模的意义与应用

百分等级是一个地位量数，能够反映某个测验分数在一个次数分布中德相对低位。其是把学生的原始分数放在该学生所在群体的成绩中进行比较，以确定该学生在群体中的相对地位之高低。

百分等级常模是基于某个常模团体，为某种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类型。

百分等级常模意义直观，容易理解，便于解释，在能力测验和学业测验中得到广泛的应用。

二、百分等级常模的建立方法

1. 基于未归类数据建立百分等级常模的方法
2. 基于分组归类数据建立百分等级常模的方法

第三节 标准分数常模及建立方法

一、标准分数的基本定义

标准分数是以标准差为单位表示测验成绩与平均分数之间的距离。

二、标准分数常模的建立方法

是以常模团体在某一测验上的实测数据为基础，把原始分数转换成基本标准分数 Z 或转换到更大的标准分数 T 量表上，能揭示每个测验分数在常模团体测验分数中的相对地位的一种组内常模。

第四节 标准分数在我国高考中的试点应用

一、在高考中使用原始分数的局限性

1. 原始分数未能反映考试分数相对于团体的位置信息
2. 不同科目或同一科目不同次考试之间分数可比性较差
3. 各科标准差不一影响对科目权重的设计

第七章 教育测验的标准及其建立方法

第一节 教育测验标准的含义

教育测验标准指教育测验的成就标准，其对于编制标准参照以及解释标准参照测验的分数具有重要的意义。

第四节 标准参照测验的定量标准——合格分数线

一、基于题目分类下的经验判断方法

是利用任课教师对成就测验标准的“格式塔”心理现象和教师具备的考评经验来分析测验题目，并作出分类与判断的思想方法。

二、基于题目双基度指标下的经验评判法

标准参照测验题目的双基度指的是题目的考察内容和测量目标对体现特定学科“基础知识，基础技能”双基教学要求的適切性程度。题目的双基度可在 0-1 之间取值。

三、基于题目“目标-难度”层次评估法

基于题目“目标-难度”层次评估来确定一个测验的合格线时，其主要思路是：根据对题目考查目标层次的认识以及对题目难易程度作出的经验判断，把整个测验卷中所有题目按照两个特征进行双向分类，然后，由科任教师给每一类题目指派一个临界概率。这个临界概率是，设想一个尚可接受的最低能力的学生能够正确回答该类题目的可能性大小，或者说这一临界概率是对每一位合格的学生至少应答对该类题目的比例大小（同概率值大小相等）。

四、基于已知“掌握-未掌握”效标组的统计分析方法

第八章 学生课业发展的测量与评价

第一节 课业考评改革的基本认识与目标

一、课业考评的主要作用

1. 合理的课业考评制度为学生发展提供较明确的目标和努力的方向
2. 合理的课业考评制定将有助于评价学生的发展、进步，从而对教与学双方活动起着重要的控制、调节和促进等作用
3. 课业考评为学生心理发展和学习进步创造必要的背景和空间，诱发学生的学习动机和自主发展的动力
4. 课业考评在中小學生个体社会化进程中起着控制、调节、促进和加速作用

四、课业考评改革的主要目标

1. 要实现考试观向发展性评价观的转变

发展性评价观，本质上是秉持一种“以人为本、以学生为中心、以促进学生全面发展为根本宗旨”的现代教育评价理念及方法体系。

2. 要努力实现从重视“掌握性结果”的评价转变到既重视“掌握性结果”又重视“发展性结果”的评价

掌握性结果往往与课程中那些明确表述的知识、能力和技能有关，其容易表述，容易考评，且通常可借助于纸笔测验加以评定和检查。而发展性结果与那些很难全部达到的、较难用一般的纸笔测验来检测的且富有个性化的行为目标有关，如学生思维品质、创造力等。

3. 内容要从认知领域转变到涵盖学习结果的更广泛的教育目标领域上来

课业考评要采用多种方法、多种形式，全面落实教育目标，注重对学生在德、智、体、美、劳等多个方面进行评价。

4. 要努力实现从表征性分数机制到实质性内容机制的转变

表征性分数机制以考试分数作为追求的目标和决策的依据，把学生在一个学期或一门课程中丰富多彩的、生动活泼的学习行为，抽象概括成一个考试分数；实质性内容机制则从定量考评和定性考评相结合出发，强化课业考评的教育与发展功能，从教学内容和教学目标两个维度，采用多种教育评价技术，对学生的课业进展情况作出负责任的实质性的评价。

实质性课业考评内容机制的特点：

1. 强调教育目标的整体性与广泛性
2. 重视对课程学业目标的确定与落实
3. 强化教育性功能，淡化区分与选拔功能
4. 重视用多元方法对学业的课业发展进行评估，重视过程评价和结果评价相结合
5. 强调学生个性发展与教育共性要求的统一

第二节 学生课业发展的主要内容及参照点体系

一、学生课业发展的主要内容

1. 学生的个体一般性发展
2. 学科理论知识和学科能力或技能的发展
3. 思维技能与品质的发展
4. 研究与学习技能的发展
5. 创新精神与实践能力的发展
6. 态度、观念与兴趣的发展
7. 欣赏与审美的发展
8. 适应与习惯的发展
9. 学生体育技能与素质的发展
10. 个体独特性的发展

第三节 评价学业课业发展进步的主要方法

一、客观题评价法与主观题评价法

二、表现性测验评价法

1. 定义：指客观测验以外的一类以行动、作品、表演、展示、操作、写作、制作档案资料等更真实的表现来展示学生口头表现力、文字表现能力、思维思考力和随机应变力等。

2. 方式：（1）口头测验（2）论文题测验（3）实验技能教学考试评价

三、评定量表评价法

在教育与心理测量方法中，评定量表是用来量化观察所得印象的一种测量工具。其既可用于现场观察的直接记录与评价，也可用于经过较长时间的纵向观察以印象为基础的综合评价。评定量表的使用，具有收集接近客观实际情况资料的功能，尤其适用于对学生表现性行为或作业的评价，是表现性评价的常用方法之一。

1. 评定量表的形式：（1）数字等级评定量表（2）图示等级评定量表

2. 等级评定量表的使用：（1）适应于过程评价（2）适用于成果评价（3）适用于对学生诸方面发展的终结性评价

四、档案袋评价法

1. 定义：是“教师依据教学目标与计划，请学生持续一段时间主动收集、组织与省思学习成果的档案，以评定其努力、进步、成长情形”的一种评价方法。

2. 类型 档案袋评价有不同的类型。按内容重点可把档案袋分成成果型、过程型和综合型；按生成结构可把档案袋分成结构型、半结构型和非结构型；按用途分可把档案袋分成评价型、展示型和反思学习型三种类型。

五、动态评价法

动态评价有两层含义：一是跨越多个时间点观察评估学生的进步与改变情形，了解学生动态认知历程与认知能力变化的特点和潜能。二是评价者与被评价者之间产生大量的互动，强调评价与教学相结合，实施个体化的诊断评价与教学补救。

第四节 成绩评定等级制的原理与应用

一、绝对评价等级制及应用

1. 定义：按照既定的课业标准把学生的学习成绩与表现规定为若干等级，即为绝对评价等级制。

2. 常用方法（1）教师综合评价法（2）测量结果转换法（3）核心内容参照法

二、相对评价等级制及应用

1. 五分等级制
2. 十分等级制

第九章 学生智能发展的测量与评价

第一节 学生认知过程基本能力发展的测量与评价

一、观察能力发展的测量与评价

1. 观察能力的发展

(1) 观察的目的性 (2) 观察的持久性 (3) 观察的精确性 (4) 观察的概括性

2. 观察能力的测量与评价方法

(1) 实验室评价法 即结合学校有关课程实验, 利用实验课教学或实验室有关设备, 设计实验观察, 通过实际观察活动来评价学生的观察能力。

(2) 研究报告评价法 即通过对学生的课程作业、科学研究报告、实验室实验报告等来分析、判断与评价学生的观察能力。

(3) 教师观察评价法 根据教师对学生一段较长时间的观察了解, 参照上述观察活动的目的性、持久性、精确性与概括性四个一级指标, 适度分解二级指标, 可设计一份“学生观察能力评价表”, 对学生进行观察能力的评价。

(4) 测验评价法 要客观地测量与评价学生的观察能力及其特点, 采用专门的测验去评价更有效。

二、记忆能力发展的测量与评价

1. 文字记忆测验

2. 非文字记忆测验

3. 综合性成套记忆测验

三、思维能力发展的测量与评价

方法: 语言文字推理测验、图像推理测验、数学符号推理测验

四、想象能力发展的测量与评价

方法: 作品分析评价法、心理测验评价法

第二节 学生智力发展的测量与评价

二、智力结构的理论及发展

1. 斯皮尔曼的 G 因素说

英国心理与统计学家斯皮尔曼认为人的能力由一般因素(G)和特殊因素(S)

两部分构成。一般因素 G 是笼统的综合的抽象能力，为完成所有测验所要求；特殊因素 S 是因测验不同而有所不同，可表示为 S_1 、 S_2 、 S_3 等等。

2. 桑代克的特殊因素说

美国心理学家桑代克认为智力是由许多成分组成的，不同的活动包含不同的成分组合；智力是由多种高度特殊化的因素联合起来的，智力是可分解的。

3. 塞斯顿的群因素说

美国心理学家塞斯顿采用因素分析法，根据 56 个测验之间的相关系数矩阵，抽取出七种首要的独立性较强的公因素，分别是：空间知觉因素 S ；计算因素 N ；言语理解因素 V ；词的流畅性因素 W ；记忆因素 M ；知觉速度 P ；推理因素 R 。

4. 弗农的智力层次结构模型

英国心理学家弗农提出关于智力的层次结构模型，该模型可用逻辑树加以表示。在层次结构模型中，他把斯皮尔曼的 G 因素放在最高层；第二层是两大因素群，包括“言语-教育因素”和“空间-机械因素”，这两个第二层次的主因素又依次包括多种小因素群，构成第三层次。每个第三层次的小因素，按照弗农的观点又可分解成更小的特殊因素，它们构成该层次结构模型中的第四层次。

5. 吉尔福特的智力三维结构模型

美国著名心理学家吉尔福特从操作、内容和活动的成果（产品）三个维度建立一个立体模型，用以描述智力结构。其中，操作是指个体对于原始材料刺激物的处理方式，包括认知、记忆、发散思维、辐合思维和评价 5 种。内容是指引起个体心智活动的各种刺激物，吉尔福特把它区分为图形的、符号的、语义的和行为共 4 种。产品，包括单元、门类、关系、系统、转换和含蓄 6 种。因而三个变项可组成 $120(4 \times 5 \times 6)$ 种较单独的智力因素。

6. 科特尔的流体智力和晶体智力理论

流体智力：是洞察复杂关系的能力，是在不同的环境和条件下都要显示出来的一般的普遍的能力（多依赖于遗传）；晶体智力是获得知识与技能的有效结合之能力（多依赖于环境）。

7. 加德纳的智力结构理论

确定了人的七种智力：言语智力、数理逻辑智力、空间智力、身体动觉智力、音乐智力、人际智力、自省智力。后加入符合“选择标准”的第八种智力，即自然智力。

8. 斯滕伯格的成功智力理论

斯滕伯格认为成功智力包括分析性智力、创造性智力和实践性智力，这三个方面构成一个有机的整体，只有在分析、创造和实践能力三方面协调与平衡时才最有效。

三、经典智力发展的评价方法

1. 韦氏智力量表

2. 瑞文标准推理测验

是在北京师范大学心理学系张厚粲教授的主持下，根据英国心理学家瑞文1938年设计的“标准图形渐近测验”修订的一种非文字智力测验。

内容：包括A、B、C、D、E五组，每组12题，共60题，适用于6岁以上的儿童和成人。各组测验的试题皆依难度从易到难渐进排列。

优点：（1）适用年龄范围宽，测验对象不受文化、种族与语言的限制，并且可用于一些生理缺陷者。（2）测验可个别进行，也可团体实施，使用方便，省时省力。（3）结果解释直观简单，测验具有较高的信度与效度。

3. 中小學生团体智力筛选测验

第三节 学生创新能力的测量与评价

二、创新能力的评估方法

1. 观察评估

2. 通过研究性学习加以评估

3. 科技作品分析法

4. 评定量表评价法

第四节 学生能力倾向发展的测量与评价

一、学习能力倾向测验

旨在测量一般的学习能力和潜能，即是否具备较好的学习与研究的潜在能力，而不是已在学校中学到了多少知识。

二、职业能力倾向测验

职业能力倾向测验是通过测量人的某种职业活动潜能，从而预测人在一定职业领域取得成功可能性的心理测验。

第五节 学生多元智力发展的测验与评估

二、发展多元智力的学校教学改革

主要包括：

1. 个性化的学校教育

- 2.学校教育目的的强调理解并学以致用
- 3.通过多元智力渠道进行教学
- 4.重视师徒传授制

第十章 学生人格心理发展的测量与评价

第一节 人格心理测量与评价的意义

一、人格的定义

指人与社会性联系最为密切的心理特质的总和。

二、人格的基本内容

人格有不同的侧面，通常包括：气质、性格、适应性、兴趣、态度等。

第二节 人格心理测量与评价的方法

一、人格心理测量与评价的方法简介

1.自陈量表法

也称问卷测验法，多是以被试提供关于自己人格特征的自我报告的形式出现，故称为自陈量表，是人格测验最常用的方式。

2.评定量表法（见第八章）

3.情境测验法

是把被试置于特定情境中以观察其行为反应，评价其人格特征的方法。

4.投射测验法

通过间接的方法来了解人们对某个事物的态度或内心世界，通常向被试提供一些未经组织的刺激情境，让被试在不受限制的情况下，自由地表达自己的反应，如一些模糊的人形或墨迹图等，被试在自由问答时不知不觉地把自己的思想、情感、态度等人格特征泄露出来，通过分析被试的反应结果来推断其的人格类型。

第十一章 学生思想品德发展的测量与评价

第一节 学生思想品德发展概述

二、学生思想品德发展的形成与发展

1.学生思想品德的形成

道德的形成与发展有一个由他律到自律，再由自律到自由的过程。在他律阶段，外在道德规范和制度伦理的作用尤为重要，制度规范、示范、榜样、奖惩等对人们的行为选择和习惯的养成起到主要作用。在自律阶段，个体的德行或良知成为道德及其发挥作用的核心。

2. 学生思想品德的发展

个体道德发展一般包括道德认知、道德情感、道德动机、道德意志和道德行为的发展，并且表现出年龄阶段性特征。

认知心理学派的道德发展阶段论：

20世纪20年代末，皮亚杰将道德判断的思维结构分为四个阶段：一是自我中心阶段（大约2~5岁）；二是权威阶段（大约6~8、9岁）；三是可逆性阶段（大约8、9岁至10岁）；四是公正阶段（大约11~12岁）

20世纪50年代末，柯尔伯格总结出儿童道德发展的三水平六阶段，后来对儿童道德发展阶段有了更为广泛适应性的研究结论，即：

阶段零：前道德阶段，既不理解规则，也不能用规则和权威判断好坏

前习俗水平：儿童能区分文化中的规则和好坏，懂得是非名称，但其根据行为对身体上的或快感上的后果来解释好坏。分两个阶段，阶段1：以惩罚与服从为定向；阶段2：以工具性的相对主义为定向

习俗水平：按集体所期望人们去做的行事就被认为它本身就有价值，不管其所产生的后果如何。分两个阶段，阶段3：“好孩子”定向；阶段4：以法律与秩序为定向，行为是服从权威。

后习俗水平：儿童努力摆脱掌握原则的集体或权威。阶段5：以法定的社会契约为定向，带有功利主义色彩；阶段6：以普通的伦理原则为定向，根据良心作出的决定是正确的。

第二节 学生思想品德测量与评价的意义

二、学生思想品德测量与评价的意义

1. 促进德育目标的实现和质量的提高
2. 思想品德测量与评价是德育工作科学化的保证

思想品德测量与评价应坚持定性和定量相结合的原则。

3. 思想品德测量与评价有利于发挥评价的教育功能
4. 思想品德测量与评价有助于学校德育改革

三、学生思想品德测量与评价的原则

1. 方向性原则
2. 教育性原则
3. 客观性原则
4. 有效性原则

第三节 学生思想品德测量与评价的内容及方法

一、学生思想品德测量与评价的一般内容

1. 政治素质测评
2. 思想素质测评
3. 道德素质测评
4. 法纪素质测评
5. 心理素质测评

二、学生思想品德测量与评价的方法

1. 操行评语法
2. 操行加减评分法
3. 考试考核测评法
4. 积分测评法
5. 模糊综合评判法
6. 品德情境测评法
7. OSL 品德测评法

第十二章 学生体育发展的测量与评价

第一节 学生体育发展的目标及其概述

一、学生体育发展的一般目标

学校教育工作者应当充分认识体育的重要性，理解卢梭的名言的深刻含义：教育最大的秘诀是使身体锻炼和思想锻炼互相调剂。

学生体育发展的一般目标，可从身体领域发展、体育认知领域发展、体育情感领域发展和体育动作技能领域发展四个维度分析。

第三节 学生体能发展的测量与评价

一、学生体能发展的测量与评价原则

1. 客观性原则
2. 可靠性原则
3. 有效性原则
4. 实用性原则

第十三章 现代教育测量与评价的发展趋势

第一节 经典测量理论的特色

一、经典测量理论的优点

- 1.以弱假设为基础，这些假设条件容易被绝大多数测验数据资料所满足
- 2.统计分析方法在计算上较为简单，意义上直观明了，易被教育工作者理解和掌握，应用上具普遍性
- 3.我国教育工作者在使用过程中和实践中不断形成和完善了一套适合国情的具体应用方法和应用原则

二、经典测量理论的主要局限性

- 1.统计分析方法得到的各项指标依赖于特定的被试样本
- 2.题目参数和考生得分是在不同的基础上分别求得，考生能力的估计会由于测验的改变而改变
- 3.严格的平行测验和误差相等两个假设条件在实际操作中难以实现

第二节 现代测验理论的优势

一、项目反应理论的主要优点

1. 题目难易度的估计不因样本不同而不同
- 2.考生能力的估计不因测验的改变而改变
- 3.测量误差的估计因考生程度不同而不同
- 4.为多种形式测验的实施提供了更为完美的理论和方法

二、项目反应理论的主要局限性

1. 对数学模型与实测数据的拟合要求较高
- 2.项目反应理论是单维性假设，是对任何数学模型的共同假设

三、经典测量理论与现代测验理论比较

- 1.参数稳定性
- 2.能力的比较
- 3.平行复本难以实现
- 4.缺乏预测力
- 5.测量标准误