

普通高等院校“十二五”立项教材

教育测量 与评价方法

JIAOYUCELIANGYUPINGJIAFANGFA

主编◎张洪秀

 吉林大学出版社

普通高等院校“十二五”立项教材

教育测量与评价方法

主 编 张洪秀

副主编 王 威 张园园 李俊刚

 吉林大学出版社

图书在版编目(CIP)数据

教育测量与评价方法 / 张洪秀主编. —— 长春: 吉林大学出版社, 2014. 7

ISBN 978-7-5677-1937-8

I. ①教… II. ①张… III. ①教育测验—教材②教育评估—教材 IV. ①G449

中国版本图书馆 CIP 数据核字(2014)第 163576 号

书 名: 教育测量与评价方法
作 者: 张洪秀 主编

责任编辑、责任校对: 李伟华
吉林大学出版社出版、发行
印刷开本: 787×1092 毫米 1/16
印张: 15 字数: 320 千字
ISBN 978-7-5677-1937-8

封面设计: 可可工作室
北京明兴印务有限公司 印刷
2014 年 7 月 第 1 版
2014 年 7 月 第 1 次印刷
定价: 32.00 元

版权所有 翻印必究

社址: 长春市明德路 501 号 邮编: 130021

发行部电话: 0431-89580026/28/29

网址: <http://www.jlup.com.cn>

E-mail: jlup@mail.jlu.edu.cn

前 言

教育测量与评价是现代教育活动中不可缺少的重要组成部分,它对教育教学活动的开展起着导向和调控的作用,已经日益成为当代教育科学研究和教育实践的重要领域。为了适应教育测量与评价学科的发展,更为了提高广大教育理论工作者和教育实践工作者的现实需要,我们在参阅了大量的国内外相关学科的研究成果的基础上,编写了这本适合于高等院校、教育管理机构、教师培训机构等人员需要的教材,希望能为广大读者提供关于教育测量与评价的系统实用、通俗易懂的入门教材。

为了让读者从中获得教育测量与评价这一学科较为系统的理论与实践知识,本书在编写的过程中力求体现以下特点:第一,内容全面。本书由教育测量与评价的历史发展说起,阐述了该学科的历史发展脉络与特点,接着从教育测量、教育评价两部分分别对其进行了详细的介绍,内容细致详实,涵盖了该学科领域的核心概念,并介绍了当前较为前沿的理论与实践研究。第二,注重基础。本书立足于学科的基本理论与实践的介绍,强调对每一个知识点的基本理念、基本观点、基本用途的介绍,努力做到夯实学习者的基础。第三,讲求实用。本书在编写过程中,针对于各层次的教育测量与评价的使用人员不同特点,从各方面进行举例,力图将理论如何转化为实践的思路与方法体现出来,帮助学习者活学活用,灵活掌握,做到“学以致用”。第四,受众面广。本书对于高等院校的师范类本科生掌握这门工具性的学科有较好的帮助,无论从理论,或是实践方面均能获得提升;对于中学、小学、幼儿园的管理人员、一线教师等在教育测量与评价方面知识的获得、能力的提升也有很好的帮助。第五,方便使用。本书在每章结尾部分均设置了“本章小结”栏目,有助于学习者提纲挈领把握每章的知识要点,方便复习;每章最后还设置了“思考与练习”栏目,对于学习者每章自测,随时掌握学习效果都有一定好处。

本书由张洪秀担任主编,主要负责整个框架结构和编写体例的写作工作,以及全书的统稿和定稿工作。在教材的编写和修改过程中,李俊刚给予了很大帮助,在此表示感谢!各章节分工如下:第一章、第二章由张园园(吉林华桥外国语学院)执笔;第三章由李俊刚(佳木斯大学)执笔;第四章、第六章、第八章、第九章(其中第四章第三节由李俊刚执笔)由张洪秀(省佳木斯大学)执笔;第五章、第七章由王威(哈尔滨工程大学)执笔。

本书在写作过程中参阅了大量同行的著作和论文,也引用了他们的一些相关数据材料,对此我们都尽可能详细地作了说明,难免有疏漏之处,敬请谅解,并表示诚挚的谢意!对于在写作过程中给予热心帮助和支持的同事和朋友表示深深的谢意!同时,对于吉林大学出版社的大力支持也给予最衷心的感谢!由于编者的学识水平有限,在写作的过程中难免会有不足之处,敬请广大读者为我们提出宝贵的意见!

编 者
2014年3月



目 录

第一章 教育测量与评价概述	(1)
第一节 教育测量与评价的基本问题	(1)
第二节 教育测量与评价的类型	(12)
第三节 教育测量与评价的功能	(22)
第二章 教育测量与评价的历史发展	(27)
第一节 西方教育测量和评价的发展史	(27)
第二节 我国教育测量与评价的发展史	(35)
第三章 教育测量的质量指标	(43)
第一节 经典测量理论与教育测量的误差	(43)
第二节 教育测量的信度	(47)
第三节 教育测量的效度	(59)
第四节 教育测量的项目分析	(77)
第五节 教育测量分数的解释	(84)
第四章 教育测验编制的原理与方法	(91)
第一节 教育测验编制的原则与步骤	(91)
第二节 教育测验题目的种类与编制技巧	(103)
第三节 教育测验的施测	(115)
第五章 教育评价的一般过程	(118)
第一节 教育评价的准备	(118)
第二节 教育评价的实施	(126)
第三节 教育评价结果的处理与反馈	(130)
第六章 教育评价的指标体系	(139)
第一节 教育评价指标体系概述	(139)
第二节 教育评价指标体系的设计程序和方法	(145)
第七章 教育评价信息的收集方法	(161)

第一节	观察法	(161)
第二节	测验法	(168)
第三节	问卷法	(169)
第四节	访谈法	(175)
第五节	文献法	(181)
第八章	教育评价信息的处理方法	(187)
第一节	定量分析方法	(187)
第二节	定性分析方法	(202)
第九章	教育评价的心理调控	(211)
第一节	教育评价心理调控的意义	(211)
第二节	评价者的心理现象及调控	(215)
第三节	被评价者的心理现象及调控	(224)
参考文献	(233)



第一章 教育测量与评价概述

我们知道，教育研究的一个重要目的是不断地改进教学，努力提高教学质量，毫无疑问在教学研究中离不开对教学进行评价和测量。通过教学评价与测量，对教学质量进行测量与评价，从而对教学改革进行全面的评估，总结出成功的经验，提出进一步研究的问题，不断地推动教学研究。由此可见，对教学进行科学的测量与评价是教学研究的重要环节。如何才能对教学进行科学的测量与评价呢？如何提高教学测量与教学评价的效率呢？

要对教学进行科学的测量与评价，首先必须确立科学的测量与评价的理论体系，这个理论体系包括：教学测量的学科性质和地位、教学测量的基本原理及基本概念、教学评价的价值观念、教学评价的本质、教学评价的功能、教学评价的范围与方式、教学评价遵循的原则和指标体系等。

在本章，我们将讨论教育测量与评价的若干基本概念和基本问题，以便为学习以后各章的具体内容提供一个基本的框架。

第一节 教育测量与评价的基本问题

一、教育测量的基本概念

教育测量学是对教育现象进行定量测定的一门科学。研究如何利用测量理论和方法来分析教育活动，分属于教育科学层次的方法论学科。是统计学、心理学、测量学的基本原理和方法在教育领域中的应用。下面主要介绍教育测量的涵义、性质和作用，并以辩证唯物主义和历史唯物主义作为其基本观点，将教育科学的基本原则和教学测量的实践相结合，注意吸收国内外与教育测量学相关的学术著作和权威教材的内容，理论与实践相结合。

教育测量学与教育统计学相互作用，联系密切。首先，教育统计学所加工的原料是教育测量所提供的数据，教育测量本身的效度（validity）和信度（reliability）是由数据真实情况与可靠程度决定的，关于教育测量中的项目分析（item analysis）以及效度和信度的确定，又必须运用统计学中的相关系数和因素分析。其次，教育测量学必须运用统计学知识来编制各种量表及确定各种常模（norm），如百分位数、平均数、标准分数、标准差、正态分布等。最后，教育测量的理论和方法又是教育评价学的基础和教育科学实证研究的基本方法。

教育测量学在教育科学体系中占据相当重要的一席之地。教育测量在人才选拔、因材施教、教育评价、教育科学的实证研究和教育教学改革的实践研究中扮演着工具的角色。教育测量是测量的一种，要理解教育测量的概念，首先必须了解测量的概念及其要素。

（一）什么是测量

1. 测量的含义

测量（Measurement）通常是指根据一定法则给事物特性指派一定的数字的过程，它是人们认识客观世界的有效方法。

客观世界的事物名目繁多，特性迥异，要确定某一事物的某一特性，就必须依据一定的规则或方法进行测定，并将测定的结果用数字表示出来，在人们日常生活中，测量是被普遍运用的方法。例如，人们用天平、磅秤测定物体的质量，用米尺测量物体的长度，用量筒测量液体的体积，用温度计测定物体的温度等，这些都是人们日常生活中所进行的测量。这些测量并非随意定夺，而是按照一定的原理和规则，使用一定的量具进行的。人们正是凭借着测量手段，对客观事物进行数量的测定，从而更好地认识事物和对事物进行分析比较的。

美国测量学家斯蒂文斯（S. S. Stevens）曾对测量做以下定义，“就其广义来讲，测量是按照法则给事物指派数字”，这个定义简要地说出了测量的基本性质，它主要包括以下三个要素：参照点、单位、量表。

2. 测量的要素

任何测量都一般应具备三大要素：参照点、单位、量表。

（1）参照点

参照点：是计算事物数量的起点，又称零点。要确定事物的量，必须有一个计算的起点，这个起点就叫做参照点。参照点不统一，量数所代表的意义就不同，测量结果就无法进行比较。参照点有两种类型：一种是绝对的零点，如测量轻重、长短等都以零点为参照点，这个零点的意义为“无”，表示什么都测不到。另一种是人为指定的参照点，即相对零点。例如以海平面为测量陆地高度的起点，以冰水混合物为测量温度的起点。教育测量中较多地使用相对零点。这种相对零点的参照点有一个很大的限制，就是只能进行加减运算，不能进行乘除运算，且从该起点起计算的数值不能以“倍数”的方式解释。例如，在某次测验中得零分的学生未必对该测验的内容有关的知识一点也不懂。但这种人为确定的参照点只能表示差异的大小，而不能表示倍数。又如，A、B两个学生，在语文测验中，A得了20分，B得了40分，我们此时只能说B比A的成绩高20分，而不能说B是A的2倍。

（2）单位

单位：是计算事物的标准量的名称。例如，测量长度以“米”为单位，测量质量以“千克”为单位。单位是测量的基本要求，没有单位就无法进行测量，没有单位，数量的多少、大小就无法表示。单位的种类、名称繁多，即使是测量同一事物，也可以用多种不同单位。



理想的单位必须具备两个条件：一是有确定的意义，即同一单位在大家看来意义相同，也就是所有人的理解意义要相同，不允许出现不同的解释；二是要有相等的价值，即相邻两个单位之间的差别总是相等的。也就是说，第一单位与第二单位间的距离等于第二单位与第三单位间的距离。长度、重量等物理测量的单位符合这两个条件，而教育测量所用的单位则不等值。例如，有 A、B、C、D 四个学生，在某次测验中，他们的成绩分别为 20 分、30 分、50 分、60 分。那么，我们能否说 A、B 的学业水平之差和 C、D 的学业水平之差相等呢？对此，我们无法肯定。因为各测题的难度不同，所花费的心理能量也是不等的，也就是说各测题的分数单位是不等的。

(3) 量表

测量时将测量对象放在一个有参照点的连续体上，以便得到一个测量值，这个连续体称为量表 (measurement scale)。如尺子是度量长短的量具，天平是权衡重量的量具。即量表是根据测量目的所设计的测试项目和赋值规则。教育测量常用的工具是试卷，而试卷由测题构成，故我们必须高度重视命题的研究，提高试题质量，这样才能保证测验的可靠性和有效性。

由于事物的属性不同，制定的规则不同，所以不仅用数字描述事物的属性所达到的程度不同，而且用数字描述事物的属性的含义也不同。根据测量的精确程度不同，斯蒂文斯 (S. S. Stevens) 将测量从低级到高级分成四种水平，即类别量表、等级量表、等距量表和比率量表。高级量表除具有低级量表的全部特性外，还有自己独特的性质。

① 类别量表 (nominal scale)

类别量表也叫称名量表，它是量表中测量水平最低的一种。它根据事物的某一特点，对事物属性进行分类，并用数字或符号表示。它仅能区别不同类别，没有任何数量的意义，只起着标志事物的作用，因而没有序列性、等距性和可加性。如性别、国籍等。例子：男生为 1，女生为 2，喜欢语文的学生为 1，喜欢数学的学生为 2。在统计上，类别量表只能计算每个类别的次数 (频数)。适用的统计方法属于次数统计，如计算百分比或列联相关，进行 χ^2 (卡方) 检验等。

② 等级量表 (ordinal scale)

等级量表也叫顺序量表，其测量水平比类别量表高，它根据事物的某一特点，将事物属性分成等级，用数字表示。这种测量水平，不仅能区分不同类别，而且能排出等级或顺序。如胖瘦、高矮、大小、优良中差等。例子：根据学生测验的成绩排出名次，成绩最好的为 1，成绩次之为 2，再次之为 3，依次类推；又如，对于生活水平，我们可以分为四个等级：贫困、温饱、小康、富裕。等级量表反映事物的类别的差不必相同，其只具有等级性或序列性，不具有等距性，也没有可加性。它所适用的统计方法有中位数、百分数、等级相关系数、斯皮尔曼等级相关和肯德尔和谐系数以及秩次变差分析等，但不能做加、减、乘、除运算。

③ 等距量表 (interval scale)

等距量表也叫间距量表，它对事物属性的划分是等距的，即它们的单位是等值的，但没



有绝对零点。它的参照点是人为指定的，只具有相对性质。因此，此类测量所得的数据可进行加、减运算，但不能进行乘、除运算。例如：摄氏温度计就是一个典型的等距量表，摄氏20度不能说成是摄氏10度的2倍，而只能说他们两者相差10摄氏度。又如：三个儿童在智商测验中分别得分105、110、115，在智商测验分数体系中，分数差距是相等的。等距量表能更广泛地应用多种统计方法，如均数、变差、积距相关系数、T检验和F检验等。

④ 比率量表 (ratio scale)

比率量表又称等比量表，它是一种理想的量表，也是最高水平的测量。它除具有等距量表的一切特性外，还具有绝对零点。如在物理测量中，长度、重量等都是比率量表。例子：摄氏零度，这里零度并不意味着没有一点温度，而是以人定的冰点为参照标准。像学生的考试成绩，智商得零分都是相对零点。比率量表所得数值可做加、减、乘、除运算。它所适用的统计量除上述几种外，还可以使用几何平均数和相对差异量。

表 1-1 四种量表的比较

量表名称	基本特征	示例	性质
类别量表	相互排斥、具有可辨别的类别	性别量表	①区分性
等级量表	等级、位次、大于或小于	优良中差百分制	①②顺序性
等距量表	量表上单位相等	温度计	①②③等距性
比率量表	有绝对零点	尺子、天平	①②③④可加性

以上四种量表是依从低到高的次序排列的。量表的次序越高，对于描述事物的数所能允许的算术运算也就越多。后面每一种量表的性质，除包括前面各种量表的性质之外，还具有特殊的性质。人们往往把类别量表和等级量表进行的测量，认为是定性测量。把等距量表和比率量表进行的测量，认为是定量测量。

(二) 什么是教育测量

1. 教育测量的含义

1918年，美国心理学家桑代克(E. L. Thorndike)提出了“凡物之存在必有其数量”；1922年，美国测量学家麦柯尔提出了“凡有数量的东西都是可以测量的”；二千多年前，我国孟子对齐宣王说：“权，然后知轻重，度，然后知长短，物皆然，心必胜，王请度之”。人类甚至从远古的蒙昧时期就已经开始了对身边周遭事物的有意识或无意识的测量估算，人类几乎无时无刻不再算计、评估，这种算计、评估被规范化了，其实就是我们要讲的测量。测量是人们根据某种规则对客观事物进行的数量化测定。

教育测量学是研究如何测量学生的学业成绩、学习能力、兴趣爱好、思想品德以及教育上许多措施的效益等问题的一门科学。它是改进教学实现科学化教育管理和教育研究的重要工具，是教育评价方法技术的一部分。教育测量是一种间接测量，教育测量的哲学基础是它的存在性、唯一性和可测性。学习者的知识、能力、技能、态度等是客观存在的，教育教学



过程是客观存在的，因而是有数量的，是可以测量的。教育测量是根据教育目标的要求，按照一定的规则对教育活动中涉及到的学习者的知识、能力、技能、态度、教学过程的效果等加以数量化测定的过程。其主要内容包括：测量原理、项目分析、效度、信度、标准化测验、测验的设计和编制、测验的常模、测验的实施和评分、测验结果的解释等。

2. 教育测量的特点

教育测量主要测量学生的内在心理特征，它与物理测量有不同的特点，主要表现在以下四个方面：

(1) 教育测量的间接性和推断性

物理测量可以运用工具测量物体的本身。如一个物体的质量、长度、温度，可以用天平、尺子、温度计直接测量，一般采用的是直接测量的方法，来认识物体的物理属性。而教育测量所要测量的对象是人的知识、智力、才能，是人的大脑活动的表现，属于行为属性，不能直接测量。我们只能依据人的外显行为，间接的测量人的心理活动的特点和水平。我们只能通过学生对测验题目的反应和一些行为表现，运用推理、判断的方法，来间接的测量出他的知识水平、智力高低和品德好坏。因此，教育测量比物理测量要复杂得多。在教育测量中，除了对学生身体的素质和发育的测量（如身高、体重、视力、听力等）可以直接测量外，其余则是间接测量。

(2) 教育测量的尺度具有相对性

对于物理测量来说，它的标准是具体的，有形的，度量单位是绝对的。如人们用天平的砝码作为称重的标准物，用米尺作为度量长短的标准物。这种标准物看得见、摸得着；一米就是一米，一磅就是一磅，无论北京，还是上海，都是相等的；50克比45克重5克，20克比15克重5克，这两个相差的5克是相等的，因而，物理测量是比较客观的和绝对的。但是，对于教育测量来说，情况就比较复杂了。从多数学校所用传统考试方法对学生学业成绩的评定，以及学习能力、思想品德的评定来看，其测量的标准存在于测量者的头脑之中，这些标准物一般都是无形的、抽象的，因而，在很大程度上是不确定的，相对的。如甲在某校的70分并不一定比另一学校乙的80分低，因为考试的题目难易不同，评分标准不同，因而分数的价值并不相等。又如一个在测验中得80分的学生知识量并不等同于在同一测验中得40分学生知识量的两倍。因此，教育测量的度量单位一般是相对的，必须对教育测量的数据进行转换，否则，不能进行代数运算。一个学生的学业成绩好坏、智力高低和能力大小等，都是就其在所在团体的整个分数序列、行为序列中的地位来说的，其测量的分数单位，并非绝对的。为了使教育测量的分数具有可靠性、可比性，必须根据标准测量的原理和方法，编制测验量表，制定科学的评分标准和方法，并将测量所得的原始分数转换为可以直接比较的量表分数。

(3) 教育测量对象的模糊性及测量尺度的不确定性

测量是一个与选定的标准物相比较，以对被测物体赋值的过程。因此，测量所依据的标



准物对测量的质量有较大的影响。物理测量的对象是具体的、有形的、看得见、摸得着的。因此，它具有较大的客观性，可以保证其较大的精确度，甚至可以保证两次以上测量的同一性。例如，用米尺作为度量长短，测量两次以上，误差也非常很小。而教育测量的对象主要是学生的精神属性，教育测量情况比较复杂。学生的学业成绩、品德、智力等各方面情况是在不断变化、发展的，其测量的标准存在于测量者的头脑之中，这些标准物是无形的、抽象的、不能直接测量。例如，对同一被试者的测量，不同的测评者使用的尺度往往是不一致的，甚至，对同一被试者的测量，同一测评者不同时间的两次的测评结果也可能是不同的。有些主、客观因素，也会影响测量的结果，因而，在很大程度上是不确定的。其次，教育测量活动具有不可复验性，物理测量活动可以进行重复和检验，但对于教育测量来说，一般不具有这个品质。教育测量中，教师作为学生学习质量的测评者，她所使用的尺度在她的头脑中，我们难以用她使用的尺度对她的测量进行检验，因此，教育测量比物理测量要困难得多。

(4) 教育测量具有明确的目的性

任何一个测量都具有明确的目的，教育测量是为改进教育工作，提高教育质量，更好地实现教育目的服务的。即为了掌握学生在学业、智力和思想品德等方面的情况，了解教育和教学的效果，以便更合理地组织教育活动，安排教学内容，选择教学方法，贯彻因材施教，促进学生的全面发展。教育测量是通过了解、评价、诊断，从而改善教育教学，因此，测量是手段不是目的。那种不顾教育目的，为测量而测量的做法，是不符合教育测量的要求的。

表 1-2 教育测量的特征

	物理测量	教育测量
对象属性	具体而稳定	抽象而易变
方式	直接测量	间接测量（外显推断内隐）
度量单位	绝对、有意义	相对、不具备确定意义
参照点	常有绝对零点	一般没有绝对零点、用相对零点
测量结果	有良好的数学性质	没有良好的数学性质

3. 教育测量的作用

学习教育测量学对于提高教育、教学质量，提高教学管理水平，进行教育研究都具有十分重要的作用。

(1) 教育测量有利于提高教育、教学水平

教育和教学的最后一个环节就是对成果的检验，同时在教育和教学过程中，也需要随时了解教育和教学情况，诊断学生学习中存在的问题，实行因材施教，所以这些方面都需要用到教育测量的理论和方法。教育测量能为教师提供教学信息反馈，它可以帮助教师准确地发现学生掌握知识的程度，评定学生的学习进步，有针对性地采取补救教学等提供有效信息，从而促进教学效果的不断提高。同时，教学测量可以促进教师对教学目标理解，教师只有对



教学内容、教学目标有深刻的理解，才能在教学中把握正确的方向，编制出高质量的考题，这对提高教学质量起着巨大的作用。教学测量还能激发学生的学习动机，编制较好的、及时的测验，能够更有效地激发学生的学习兴趣，提高学习效率。因此，正确的利用教育测量，有利于提高教育、教学水平。

(2) 教育测量有利于提高教育管理水乎

教育测量是根据一定的目的，编制客观的标准化的量表，进行广泛的测量，做出科学的分析和评价，使我们摆脱主观经验型的笼统报告，从而对教学进行合理评价，评价结果可为教育行政部门指导教学、进行教育决策提供一定的依据。对于合理使用教育经费，科学选报人才，进行教育规划、决定教育政策等起着重要的作用，是教育管理不可缺少的手段。加强教育测量，能提高教育管理的科学性，能有效地改变长期以来形成的凭主观经验衡量教育效果的做法。

(3) 教育测量有利于发展教育研究方法

教育测量的研究是教育研究的重要组成部分，没有经过教育测量检验的任何一种新的教育理论、新的教材、新的教学方法，都无法科学地评定其价值，也无法判断其是否正确。对于学校教育质量的评价考试改革问题等在教育工作中的诸多问题的研究，同样离不开教育测量，离开了教育测量就难以得出客观的精确的依据。教育测量的理论与方法的丰富和发展必然会充实和发展教育研究的方法，很多教育问题的研究要直接运用教育测量的方法。教育测量已是教育科学研究不可缺少的工具。

二、教育评价的基本概念

(一) 什么是评价

通过测量，我们对事物的属性可以用教学的方法来进行描写，但这并不是我们对问题研究的最终目的，我们还应该进一步对该事物进行评价。

所谓评价，就是对测量的结果进行价值判定。评价从辞源上可以解释为“评定价值”，“是一种价值判断的活动，是对客体满足主体需求程度的判断”。从管理学的角度来看评价是“描述、收集和分析有用的、客观的资料并将这些资料转变成资讯，提供给决策者作为主观价值判断的历程。其目的主要在于提供改进之道，并进而协助决策者选择合理的行动方案”。由此我们可以看出评价的本质属性是价值判断，并借此收集信息、提供决策、完善工作，以此实现价值。

(二) 什么是教育评价

1. 教育评价的含义

早在 20 世纪三四十年代，在美国教育家泰勒主持的“八年研究”中被正式提出“教育评价”概念，注重效果的泰勒认为，“评价在本质上是一个确定课程与教学计划实际达到教育目



标的程度的过程”。到了 20 世纪六十年代，注重资料获得的美国克龙巴赫把评价广义地定义为“为获取教育活动的决策资料，对参与教育活动的各个部分的状态、机能、成果等情报进行收集、整理和提供的过程”。他提出三种使用评价的决策：（1）教程的改革；（2）关于个体的决策；（3）行政调控。斯塔费尔比姆在 1969 年提出：“评价是为决策提供有用信息的过程”。斯克里文和豪斯给出的定义是：“评价是一种对优缺点和价值的评估，是一种既有描述又有判断的活动”。这些定义都是从评价的功能上作出的，认为评价是为了更好地决策搜集和使用信息的过程。这种评价的界定对后来评价实践产生了深远的影响。国内教育评价专家陈玉琨认为：“教育评价是对教育活动满足社会与个体需要的程度作出判断的活动，是对教育活动现实的（已经取得的）或潜在的（还未取得，但有可能取得的）价值作出判断，以期达到教育价值增值的过程”。

分析国内外专家、学者对教育评价所给的定义和描述就会发现，尽管他们在提法和描述上有许多差异，但从不同侧面反映出教育评价具有以下特点：

第一，教育评价具有连续性和系统性。任何教育现象都是在原有的基础上发展变化起来的，因此，教育评价是一个有严格程序的、连续的系统活动过程。这个活动过程是有计划、有目的地进行的，一般要经过确定评价目标，搜索信息资料，综合分析资料，形成价值判断，指导今后行动等程序，不能仅看教育现象在某一时刻的点值。

第二，教育评价以教育目标为基准。教育目标是根据人与社会发展的需要，对教育活动的目的、方向和要求的规定，是教育活动的结果所应达到的标准、规格和状态，它是教育工作的出发点和归宿，是评价教育活动成效的依据。

第三，教育评价是以事实判断为基础的价值判断。价值判断是教育评价的本质特征，评价始终以对评价对象的功能、作用、状态给予价值判断为核心。

第四，教育评价要运用科学的方法进行整理、处理、解释和分析，便于进行准确的判断。

第五，教育评价是评价者与被评价者的双边合作活动。评价中的评价者与被评价者应该是相互协商的，尽量促使被评价者参与并取得被评者的支持，重视被评者的自我评价。做到主客体互动、评价和指导统一，才能达到改进工作、提高质量的目的。

根据上述内容，我们把教育评价初步界定为：“教育评价是根据一定的教育价值观或教育目标，运用可行的科学手段，通过系统地搜集信息、分析解释，对教育现象进行价值判断，从而为不断优化教育和教育决策提供依据的过程”。

2. 教育评价的现实意义

研究、学习教育评价理论对于教育决策、学校管理、提高教育教学质量和办学水平都有重要的意义。

第一，有利于端正办学方向，树立正确的教育观和质量观。

衡量一所学校的办学质量是看升学率和尖子生，还是看大面积提高教育质量和学生素质与适应能力，这是两种教育观和质量观的根本区别，也是确定学校办学方向的根本依据。因



此,当前从我国国情出发,运用现代教育评价理论,制定一套普通教育评价方案是端正办学方向,全面贯彻教育方针,实现教育目标的建设性措施。具体地说,改革传统的评价方法,采用和实施多指标综合评价,不仅有助于明确办学思想,端正办学方向,而且更重要的是使学校领导、教师和学生的评价观、教育观和学生观都发生了重大变化,实现了教育观念的重大转变,为实现教育目标、深化教育改革奠定了思想基础。

第二,有利于加强对教育事业的宏观指导和管理,实现管理工作的科学化。

加强对教育事业的宏观指导和监督是教育体制改革的客观需要。在实行基础教育由地方负责、分级管理的原则,权力下放之后,国家和地方教育行政部门要加强宏观管理和指导,就必须建立系统的教育评价和监督制度。因此,教育评价就成为教育领导部门和督导部门对中小学科学管理和监督的重要手段之一。

就学校内部来说,教育评价也是学校领导加强各项工作管理,实现管理目标,提高管理科学水平的有力手段。因为学校工作质量综合评价的标准体系,既是衡量学校办学质量的科学标准,也是学校进行科学管理的可靠依据。如果学校内部以自我评价的形式,使教育、教学工作与经常性、形成性评价相结合,就能在评价的动态过程中充分发挥评价的多种功能,从而达到不断完善学校管理科学化、改进各项工作、实现理想的教育目标。

第三,有利于深化教育改革,普遍提高教育教学质量。

实施教育评价与深化教育改革,在实践中是密切相关、相辅相成的。因为实施教育评价本身就是一场深刻的教育改革,它是教育改革的重要内容和组成部分,同时教育评价又是推动改革、衡量教育改革成果的重要尺度。

第四,有利于掌握信息,为教育决策科学化提供依据。

教育评价不仅是对教育现象进行价值判断的动态过程,而且也是以搜集大量评价信息为教育决策提供依据的过程。因此,在教育评价的实践中不仅应当重视评价结果的解释,更应重视掌握大量信息资料,经分析后为制定各项教育决策服务。

三、教育评价与相关概念的辩证关系

(一) 教育评价与教育测量

教育评价与教育测量是既有密切联系又有很大区分的两个不同概念。

教育测量是教育评价的初级阶段,教育评价是在测验运动的批判中发展起来的。教育评价的许多评价量表的编制要用到教育测量的方法,教育评价的信度、效度、难度、区分度的测量也离不开教育测量学。因此,教育测量是教育评价的基础,若没有教育测量,就没有可能进行教育评价的素材。从教育评价的发展历史来看,教育评价是教育测量的深化,教育评价是对教育测量结果的进一步判定,若仅有教育测量,而不进行教育评价,则无法确定该事物属性的价值如何,是优还是劣,因此,也可把教育评价的过程看作是一个总结的过程。

教育评价与教育测量是有区别的。教育测量强调数量化,即着重以数量化的标准客观地

表现某人具有的特质，它注重可观察现状的获得，并证明其数量的多少，这一切须借助于测验才能实现。而教育评价则突出价值观，它采取质、量并重，主、客观相结合，不仅要知其数量多少，更重要的是了解其发展变化过程，它对某人某事进行解释、诊断或价值判断的目的在于改进缺点和不足，促进教育、教学，以求得最佳效果。教育测量是一种单一的活动，而评价则是一个综合的活动。教育评价可采取各种方法和手段，不一定非要测验或完全依靠测验才能实现。因此，教育评价的范围之广，方法之复杂、繁多，乃是教育测量所不及的。

（二）教育评价与教育评估

教育评价与教育评估实际上是非常相近的两个概念。“教育评估”一词最早见于1985年颁布的《中共中央关于教育体制改革的决定》，1990年国家教委颁布的《普通高等学校教育评估暂行规定》中也称“教育评估”。但是1986年3月，我国政府关于第七个五年计划的报告中则正式提出“要加强教育事业的管理，逐步建立系统的教育评价和监督制度”。这里又称“教育评价”。1984年12月曾召开由10个单位组成的“教育评价课题协作组研讨会”。在此基础上于1986年在我国首先成立了全国性学术组织，称为“全国普通教育评价研究会”。1990年正式定名为“全国普通教育评价专业委员会”“教育评价”一词一直沿用至今。在我国高等学校中称为“教育评估”居多，1994年1月正式成立了全国性学术团体“中国高等教育学会高等教育评估研究会”，这里称为“教育评估”。在我国正式出版的学术著作中称为“教育评价”者占绝大多数，但也有人称为“教育评估”。比如，1987年张玉田等编著的《学校教育评价》，1988年陈漠开主编的《高等教育评价概论》，1989年瞿葆奎主编的《教育评价》。我国学者陈玉琨1987年有一本著作，名为《教育评估的理论与技术》；而1993年的另一本专著则称为《中国高等教育评价论》。1995年出版了王致和主编的《高等学校教育评估》，而同年出版的魏超群专著则称为《数学教育评价》，1996年出版了侯光文的专著《教育评价概论》，1998年出版了陶西平主编的我国第一部《教育评价辞典》。尽管不同学者在理解上略有不同，但并没有严格的区分标准，可见，教育评价和教育评估在使用中并无严格的界限。

不过，这两个概念还是有一定差别的。教育评价既有对群体、单位的价值判断，又有对个人的判断。而评估多用于对群体或单位的状态及效果的估价，如产量评估、资产评估、房地产评估、实验室建设评估、办学水平评估、课堂教学水平评估、人的心理评估等；评价相对而言是较精确的价值判断，而评估较为模糊和粗略。

（三）教育评价与教育统计

教育统计学是将数理统计学的理论和方法运用于教育领域，从数量的角度研究教育现象的数量特征及其关系的一门应用学科。就其内容来说可以分为描述统计、推断统计和多元统计分析三大部分。描述统计主要研究如何简缩数据和描述这些数据，如制成表格，画出图形，计算一些常用的统计参数，用以揭示某些方面的特征；推断统计则是通过样本所提供的信息，



对总体的某些特征进行推断、估计和预测，以揭示事物的内在规律；多元统计分析包括多元回归分析、多元方差分析、对应分析、因子分析、主成分分析、聚类分析、判别分析、典型相关分析等，主要是研究和处理多因素、多变量的教育问题。

教育评价的任务主要是研究教育活动是否遵循教育规律，判断所确定的教育目标实际上实现的程度。这些研究，作为基本工具的教育统计方法是必不可少的。尤其是比较复杂的教育活动，如何探讨和掌握教育活动规律，如何评价教育活动结果，如何科学地进行评价就更需要学习和掌握教育统计方法。反之，教育评价的理论和实践研究，又给教育统计提出了新课题，扩大了视野和新的研究领域。教育统计为教育评价提供数据处理的工具，也为评价结果的质量提供保证。

（四）教育评价与教育研究

教育评价与教育研究有很多相似之处。二者都有一定的研究课题和目标，都需要大量且全面的资料、信息、数据，并且都要对获得的资料、信息、数据进行整理归纳，都要运用定量、定性方法进行分析比较、综合处理、深入研究，最后形成总结报告。但二者又有本质上的区别，主要表现在：1. 二者的目的不同。教育评价的目的则在于通过评价来调节、控制教育过程和教育行为，进一步完善和提高教育质量，提高国民素质。而教育研究的目的在于通过研究作出新的发现，获得新的理论。2. 二者结果的意义不同。教育评价的结果则直接为人的自我完善和教育决策服务，为教育管理服务。而教育研究的结果，有的可以直接运用于实践，有的可以丰富、发展教育理论和科学技术，间接为社会实践服务。3. 二者的本质属性不同。教育评价的本质在于价值判断，在于调节和控制。教育研究的本质在于有所发明，有所发现，有所创造。因此，教育评价和教育研究是本质不同的两种教育现象。

（五）教育测量与教育考试

教育测量是对教育的属性、特征及其运动和发展规律的定量描述，主要用于对学生的精神特性进行数量化测定。即教育测量就是对学生的思想品德、学习能力、兴趣爱好、学业成绩、身体素质以及教育措施上的许多问题的数量化测定。教育测量是以考试与测验引起学生的某种行为，并对行为的结果作定量描述，这一定量描述就是教育测量的结果。

教育测验：即考试，是对人的知识、技能、能力以及某些心理特征的测量。考试评价是教育评价的核心，它是以考试为主要手段，对教育活动的过程和结果进行测定、分析、比较，并给以价值判断的过程。它由主考、被考（考生）、试卷（试题的集合及其评分标准）、考分四要素组成。考试的过程就是主考组织编制试题，通过试题作用于被考，将正确答案、评分标准与被考的反应（答案）相比较，然后给出考分的过程。

考试从宏观上可分为社会考试和教育考试两大类。教育考试又可分为校内考试和校外考试两类。根据考试方式的不同可以分为：直接考试（命题人员明确要考什么）、间接考试（迂回评估）；根据考试目的的不同可以分为：水平/能力考试（如高考）、成就考试（如期中、期