

# 教育测量与评价

JIAOYU CELIANG YU PINGJIA

◎主编 徐红



华中科技大学出版社

<http://www.hustp.com>

# 教育测量与评价

JIAOYU CELIANG YU PINGJIA

主 编 徐 红



华中科技大学出版社

<http://www.hustp.com>

中国·武汉

图书在版编目(CIP)数据

教育测量与评价/徐红主编. —武汉:华中科技大学出版社,2016.8

ISBN 978-7-5680-1156-3

I. ①教… II. ①徐… III. ①教育测验 ②教育评估 IV. ①G449

中国版本图书馆 CIP 数据核字(2015)第 200706 号

教育测量与评价

徐 红 主 编

Jiaoyu Celiang yu Pingjia

策划编辑:曾 光

责任编辑:倪 非

封面设计:孢 子

责任校对:何 欢

责任监印:朱 玢

出版发行:华中科技大学出版社(中国·武汉)

电话:(027)81321913

武汉市东湖新技术开发区华工科技园

邮编:430223

录 排:华中科技大学惠友文印中心

印 刷:武汉科源印刷设计有限公司

开 本:787mm×1092mm 1/16

印 张:21.75

字 数:540千字

版 次:2016年8月第1版第1次印刷

定 价:49.00元



华中出版

本书若有印装质量问题,请向出版社营销中心调换  
全国免费服务热线:400-6679-118 竭诚为您服务  
版权所有 侵权必究

# 前 言

“教育测量与评价”是教育测量学与教育评价学整合而成的一门综合性教育课程,具有综合性、基础性、技术性、实践性、应用性等特征。教育测量与评价既是当今教育科学研究的重要领域之一,更是当下教育管理实践的重大支柱之首。“教育测量与评价”是高校教师教育专业及教育学类专业的一门重要的专业基础课程。在高校相关专业开设本课程不仅是为了帮助学生熟练掌握教育测量与评价的原理及方法,还是为了培养学生的科学态度、创新精神和实践能力,进而培养学生研究教育测量与评价问题以及解决教育测量与评价问题的基本能力。

编写一本适合当下教师教育专业及教育学类专业大学生学习的教育测量与评价方面的教材是本教材全体编者多年的愿望与追求。本教材是全体编者在国内外已有的相关研究成果的基础上,结合自身长期的教育教学研究实践及教育教学实践活动的体悟编写而成的。在编写本教材的过程中,我们借鉴了国内外大量相关研究成果,充分考虑了当代大学生的学习特点及未来发展的需求。为了便于广大学生自我检查学习成效,我们在每章后面都列出了一定数量的思考题与练习题。

长江大学徐红老师担任本教材主编,负责全书的提纲编写及组稿工作;长江大学唐青才老师和熊猛老师担任副主编,负责全书的统稿与校对工作。本教材各章节的主要内容和具体分工为:第一章由长江大学徐红老师执笔,主要介绍教育测量与评价的相关概念、教育测量与评价的发展历程、教育测量与评价的学科地位和功能;第二章由长江大学张艳红老师执笔,主要介绍教育测量与评价的效度、信度、难度及区分度等质量特性;第三章由湖北文理学院熊德明老师执笔,主要介绍教育测验的目标、题型,以及教育测验设计与编制;第四章由上海市市西中学张雪艳老师执笔,主要介绍教育评价标准的基本概念、教育评价标准的编制依据、教育评价标准的编制方法;第五章由江西师范大学汤舒俊老师和长江大学熊猛老师执笔,主要介绍教育评价资料的收集与处理方法,以及质性评价法;第六章由湖北师范大学桂勇老师和上海市市西中学张雪艳老师执笔,主要介绍学校评价的基本理念、特色学校评价及现代学校评价标准的指标体系;第七章由湖北第二师范学院陈光春老师执笔,主要介绍学校教学工作评价、德育工作评价、体育工作评价的基本内容、一般程序及指标体系;第八章由湖北工程学院王成营老师执笔,主要介绍校长评价、教师评价及班主任评价的主要目的、基本内容、指标体系和一般方法;第九章由长江大学陈方老师执笔,主要介绍学生品德评价、学生学业成就评价及学生综合素质评价的主要内容、指标体系及基本方法;第十章由长江大学唐青才老师执笔,主要介绍课程评价的基本概念、主要内容、一般模式及发展方向。

在教材的编写过程中,我们参阅并引用了诸多专家学者的大量研究成果,也引用了他们的一些新的研究观点与研究数据,但本教材并未对此逐一详细注明,在此表示深深的歉意并请予以谅解。由于我们的学术水平及编写水平有限,书中可能还有不妥或错漏之处,敬请各位专家、同行及广大读者批评指正! 谢谢!

编 者

2016年1月

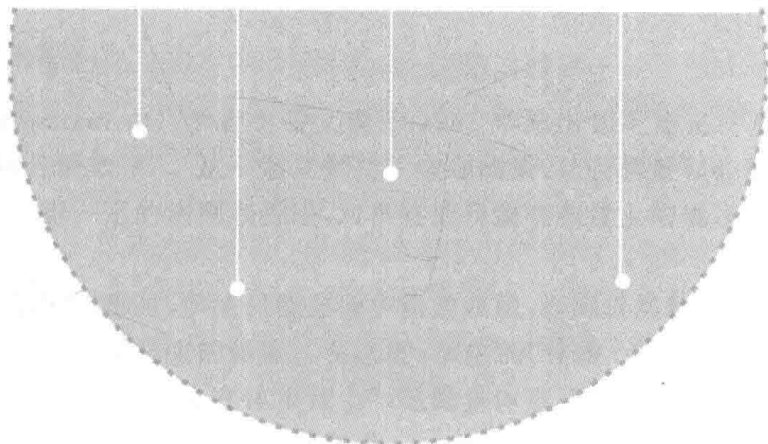
# 目 录

第一章 教育测量与评价概述	(1)
第一节 教育测量与评价的概念	(3)
第二节 教育测量与评价的发展历程	(11)
第三节 教育测量与评价的学科地位和功能	(17)
第二章 教育测量与评价的质量特性	(23)
第一节 教育测量与评价的效度	(25)
第二节 教育测量与评价的信度	(43)
第三节 测验项目的难度	(55)
第四节 测验项目的区分度	(60)
第三章 教育测验的编制	(71)
第一节 测验目标	(73)
第二节 测验题型	(80)
第三节 测验设计与试题编制	(88)
第四节 教育测验的常模	(103)
第四章 教育评价标准的编制	(109)
第一节 教育评价标准概述	(111)
第二节 编制教育评价标准的依据	(114)
第三节 编制教育评价标准的方法	(116)
第四节 编制教育评价标准的案例	(127)
第五章 教育评价的基本方法	(131)
第一节 教育评价方法的类型	(133)
第二节 收集教育评价信息的方法	(136)
第三节 处理教育评价信息的方法	(138)
第四节 质化评价法	(144)
第六章 学校评价	(155)
第一节 学校评价概述	(157)
第二节 现代学校评价的理论发展	(161)
第三节 特色学校评价	(169)
第四节 现代学校评价指标体系	(173)
第七章 教育工作评价	(189)
第一节 教学工作评价	(191)

第二节	德育工作评价	(209)
第三节	体育工作评价	(239)
第八章	教育工作人员评价	(251)
第一节	校长评价	(253)
第二节	教师评价	(263)
第三节	班主任评价	(280)
第九章	学生评价	(291)
第一节	学生评价概述	(293)
第二节	学生学业成就评价	(296)
第三节	学生综合素质评价	(304)
第十章	课程评价	(315)
第一节	课程评价概述	(317)
第二节	课程评价的模式	(326)
第三节	我国课程评价的发展方向	(331)
参考文献		(336)

# 第一章

## 教育测量与评价概述







几乎所有成功的教学和教育决策都离不开教育测量与评价,正因为如此,教育测量与评价的学科发展在教育领域乃至社会各界已被越来越多的人士所关注。教育测量与评价同时也是我国各高校教育学专业学生必修的核心专业课,在教育科学的基础学科和应用学科之间起着一种中介和桥梁作用。一方面,它是基础教育学科的深化,是从事教育基础理论研究的方法课;另一方面,它又是应用教育学科的基础,是从事教育实际应用研究的工具课。本书将在本章介绍教育测量与评价的若干基本概念和基本问题,为本书的学习者建立一个基本的框架,为学习以后各章的具体知识打下基础。

## 第一节 教育测量与评价的概念

概念是思维的逻辑起点,为方便读者了解全书内容,本节将逐一阐述教育测量、教育评价、教育测验、教育评估等核心概念,并对其相互之间的区别进行辨析。

### 一、教育测量的概念

#### (一) 测量的含义

所谓测量(measurement),学者史蒂文斯于1951年给出如下定义:“从广义而言,测量是根据法则给事物分派数字”。这一定义概括了物理测量、社会测量和心理测量的共性。具体来说,测量是指根据一定的法则和程序,对事物或现象在数量上的规定性加以描述和确定的过程。

例如,测量学生的体重时,学生只能穿极少量的衣服,赤脚自然站立在体重计上,这时体重计上所显示的数字,就是该生的体重。在这里,学生的“体重”是我们所要测量的属性,而“穿极少量的衣服,赤脚自然站立在体重计上”,是测量体重所依据的规则。体重计上所显示的“数字”,就是我们用来描述学生体重的结果。

从上述定义和举例可以看出,测量包括以下三个方面的特性。

(1) 法则。法则指测量的标准或科学原理,也就是测量的依据和准则,即根据什么原理来进行测量。例如:依据力学中的杠杆原理,或者根据作用力与反作用力相等的原理,用杠杆、磅秤等计量工具测量人的体重或事物的质量;依据热胀冷缩的原理,借助标准化的温度计测量物体的温度,等等。法则有好坏与优劣之分,使用较好的或较优的法则可以获得较为理想的测量结果,而较差的或较劣的法则只能获得不够准确的测量结果。

(2) 事物。事物指测量的对象,明确地说,即是引起研究者注意并准备加以测量的事物的属性或特征。

(3) 数字。数字是测量结果的表现形式,是区别测量与定性评价的一个标志。测量的结果必须以一定的数字形式表现出来。数字大多数时候具有量的意义,代表一定的数值。

要实现测量,必须具备以下三个基本条件,这三个基本条件即测量的三要素。

(1) 单位。意义相对明确的测量单位是测量所必需的条件之一。没有明确的单位就无法测量,而且,这一明确的单位必须是大家公认的。例如,米、秒、千克、立方米,等等。

值得强调的是,理想的单位必须具备两个条件:一是要有确切的含义,即对同一单位,所有人的理解意义要相同,不能出现不同的解释;二是要有相等的价值,即等值性,也就是相邻两个单位点之间的差别是相等的。注意:物理测量的单位是比较明确的,相同单位是等值的,如3米与2米的差等于9米与8米的差;但教育测量的单位一般是相对的,相同单位所代表的含义不一定是等值的,例如,选择题的“1分”与填空题的“1分”具有不同的含义,是不等值的。

(2) 参照点。参照点是测量或计算的起点。参照点不同,测量结果不同,且测量结果之间无法进行直接比较。参照点一般有绝对零点和相对零点两种。

绝对零点指客观存在着“0”这个数字,即“完全没有”;相对零点则是人们为了区分或分出等级,人为地指定出一个零点,即是说,即使在“0”这个位置,也不是说事物的属性是不存在的。例如:体重0千克为绝对零点,表示没有质量;考试成绩0分为相对零点,如果某学生的某门学科考试分数为0,并不表示该生对该门学科一无所知。

(3) 量表。量表也叫量具,是测量的工具。具体来说,量表是指依据某些科学原理和法则发展出来的合适的量具或制订出来的科学的测量工具。例如,气质量表、智力量表、性格量表等,又如,温度计是度量温度高低的量具,磅秤是衡量物体质量的量具。

量表具有一定的单位和参照点,不同的量表不仅具有不同的单位和不同的参照点,而且具有不同的精确度。根据量表的精确水平,可将量表从低到高分四种类型,分别为类别量表(称名量表)、顺序量表(等级量表)、等距量表(等距量表)、比率量表(比率量表)。

类别量表是最低水平的一种测量量表,它用数字代表事物或事物的类属,表示相同与不同的特性,没有数量大小的含义。例如,将学生按性别进行分类,凡男生用1表示,女生用2表示。如果既按性别分类,又按对某门学科喜欢和不喜欢两个标准进行分类,喜欢用1表示,不喜欢用0表示,于是,男生喜欢者可表示为11,男生不喜欢者可表示为10;女生喜欢者可表示为21,女生不喜欢者可以表示为20。在这里,用来描述事物的数字仅仅是代表事物的符号,其只能区分事物的类别,没有数量的大小、多少、倍数关系。也就是说,它只是具有数的同一性和区分性,其产生的数值不能用于数量化分析,不能进行数学运算,不具有等级性、等距性和等比性,因此,不能对其进行加减乘除运算。类别量表允许和适用的统计方法有比率(相对频数,即某一类的频数与总频数之比)分析、百分比分析、 $\chi^2$  相关系数。

顺序量表比类别量表要精确一些,这类量表中的数字不仅指明了事物类属,同时指明了不同类属的大小等级或具有某种属性的程度。其不仅具有区分性,还具有等级性(位次性),这些数字之间能表示事物大小的位次关系,但不具有等距性和等比性。比如,非常满意、较满意、一般满意、较不满意、很不满意。又如,将学生的口头表达能力分成甲、乙、丙三个等级,甲等用3表示,乙等用2表示,丙等用1表示,于是,对于学生口头表达能力的评定构成了 $3 > 2 > 1$ 的位次关系,但这些数字只能确定事物相等或不等的关系。在不等的情况下,只能确定大于或小于的关系,如 $3 > 2$ 、 $2 > 1$ 则 $3 > 1$ 的关系,却不能确定甲等的3比丙等的1大多少个相等的单位,因为3与2和2与1之间的差距是不相等的。因此,这类量表上的数字不能进行加减乘除的运算。顺序量表所能适用的统计方法有中位数分析、百分位数分析、等级相关系数分析、肯德尔和谐系数(多列等级相关)分析,以及符号检验、秩次检验、秩次方差分析。

等距量表更精确些,它里面的数字不仅有大小关系,而且有相等的单位和相对的零点。这种量表上的数字不仅具有区分性、等级性,还具有等距性,这种数字可以进行加减运算,但由于只是相对零点,所以不能进行乘除运算。例如,用摄氏温度计测量的温度, $9^{\circ}\text{C}$ 与 $6^{\circ}\text{C}$

之差等于 $6^{\circ}\text{C}$ 与 $3^{\circ}\text{C}$ 之差,但是这并不意味着 $9^{\circ}\text{C}$ 是 $3^{\circ}\text{C}$ 的3倍,这是因为,摄氏温度计是以冰点作为参照点的,摄氏零度并不意味着没有温度。再如,钟表上的零点,并不意味着没有时间。等距量表所能使用的统计方法有算术平均数分析、标准差分析、积差相关系数分析以及 $Z$ 检验、 $t$ 检验、 $F$ 检验等。

比率量表是最精确的一种测量量表。它既有等距的单位,又有绝对零点,不仅具有区分性、等级性、等距性,还具有等比性,因此该量表中的数字可以进行加减乘除的运算。大多数物理测量量表,如长度测量、重量测量都是比率量表,而教育测量中的量表很难达到这一水平。例如,甲生身高143厘米,乙生身高130厘米,可以说甲生比乙生高13厘米,也可以说甲生身高是乙生的1.1倍。比率量表所适用的统计方法,除了等距量表所适用的统计方法之外,还有几何平均数分析和差异系数分析。

以上四种量表是按从低到高的次序排列的。量表的次序越高,对于描述事物的数字所能允许的算术运算也就越多。本书后面涉及的每一种量表的性质,除包括前面各种量表的性质之外,还具有其特殊的性质。

教育测量的量表大多属于顺序量表,但在实际使用中却被人为地提升到等距量表或比率量表的水平。为此,我们要正确地对待和使用教育测量的结果。

## (二) 教育测量的含义

教育测量(educational measurement)是根据测量学的原理和方法对教育现象及其属性从量的规定性上予以确定和描述的过程,主要包括对学生内在的精神属性的测量,如测量学生的学习成绩、智力水平、劳动技能、品德状况、人格特征等。

这里,要注意的有以下两点。

第一,教育测量是一个数量化研究过程。在进行教育测量时,要解决的第一个问题是如何将我们所要测量到的人的各种属性进行量化表示。在教育问题中,有许多属性往往很难加以量化,有时即使加以量化,但量化水平只能停留在类别变量或顺序变量的层次,难以实现数量化分析。

第二,教育测量主要测量学生内在的精神属性,因而具有间接性。这就导致了误差甚至错误的存在。

## (三) 教育测量的本质特点

(1) 测量过程的间接性,以及测量结果的推断性。教育测量的内容通常是人的种种非物质属性,如人的知识水平、气质性格、创造能力等,这些是无法直接测量的,只能通过间接测量或推断的方式进行。

(2) 测量单位的相对性,以及测量方式的转化性。教育测量的单位一般是相对的,即仅具有相对零点,必须经过转换后才能进行相应的数学分析。

(3) 测量对象的复杂性,以及测量误差的难免性。这是因为:首先,教育测量的对象往往是人的精神属性,而人的精神属性是内在的,不能直接测量;其次,人的精神属性是多变的,比如,学生的智力、人格、劳动技能等都是发展变化的;第三,在测量过程中,一些主、客观因素也会影响测量的结果,如精神状态、注意力、天气因素等。

(4) 测量工具的多样性,以及测量数据的抽象性。在教育测量中,能够用于测量的工具

繁多,比如,按照标准化程序编制的试卷就是教育测量的一种测量工具(量表)。教育测量工具(量表)之所以具有多样性,其主要原因在于,根据多样化的教育测量对象,必须编制不同类型与不同精确度的教育测量工具。教育测量数据之所以具有抽象性,其主要原因在于,绝大多数的教育测量数据并没有特指的含义,同一测量数据在不同的环境中可以做不同的理解。例如:对某学生学业成绩获得 80 分来说,如果该生所在学生群体的学业成绩普遍低于 80 分,则表示该生的学业成绩很好;如果该生所在学生群体的学业成绩普遍高于 80 分,则表示该生的学业成绩很不好。

(5) 测量目的明确性,以及测量意图的教育性。一方面,任何教育测量的目的都是为了实现某种教育目的而来的,这使得教育测量的目的十分明确;另一方面,教育测量的意图并不在于掌握有关教育数据,而在于通过掌握有关教育数据,不断改革教育内容、教育方法、教育环境、教育条件,最终促进学生在德、智、体、美、劳等方面的全面发展。

## 二、教育评价的概念

### (一) 评价的含义

“评价”(evaluation)一词早在我国北宋时期就已出现。《宋史》中就有“市物不评价,市人知而不欺”的记载,这里的“评价”是“讨价还价”的意思。我国 2000 年出版的《辞海》(1999 年版缩影本)对“评价”一词的解释是:“评价,评论货物的价格……今亦泛指衡量人物或事物的价值。”

本书认为,评价就是衡量、评定人或事物的价值,是指评价者在掌握相关事实或资料证据的基础上,根据某种价值观对一定的人事物及其属性进行判断、衡量,或者说是衡量、判断人事物的价值。

### (二) 教育评价的含义

“教育评价”迄今尚未形成统一的定义。下面是国内外一些学者对“教育评价”(educational-evaluation)一词的几种典型解释。

(1) 泰勒(R. W. Tyler)将教育评价解释为:“评价过程在本质上是确定课程和教学大纲在实际上实现教育目标的程度的过程。”<sup>①</sup>

(2) 克龙巴赫(Cronbach)提出“评价是为决策提供信息的过程”的观点,后来有人对这一界说做了具体的描述:“教育评价是一种有系统地去寻找并搜集信息资料,以便协助决策者在诸种可行的途径(方案)中择一而行的历程。”<sup>②</sup>

(3) 格兰朗德(N. E. Gronlund)指出,教育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,其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,教育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将包括测量和非测量两种方法,用公式可以形象表达,即:

① 刘本固. 教育评价的理论与实践[M]. 杭州:浙江教育出版社,2001.

② 胡中锋. 教育测量与评价[M]. 广州:广东高等教育出版社,2006.

评价=测量(定量描述)+非测量(定性描述)+价值判断<sup>①</sup>

格兰朗德认为,评价是所有成功教学的基础。

(4) 斯塔费尔比姆(L. D. Stufflebeam)等人认为,教育评价是一种划定、获取和提供叙述性和判断性信息的过程。这些信息涉及研究对象的目标、设计、实施和影响的价值及优缺点,以便指导如何决策、满足教学效能核定的需要,并增加对研究对象的了解。<sup>②</sup>

(5) 布卢姆(B. S. Bloom)在其《教育评价》一书中指出:教育评价乃是系统收集证据用以确定学习者实际上是否发生了某些变化,确定学生个体变化的数量或程度。<sup>③</sup>

(6) 胡中锋认为,“教育评价是根据一定的价值观或价值标准,运用可行的科学手段(包括测量或非测量),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”<sup>④</sup>。

(7) 黄光杨的解释是,“所谓教育评价,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”<sup>⑤</sup>。

综合以上界说发现,“教育评价”一词包含的意思主要有如下六个方面:

第一,教育评价过程是一个价值判断过程;

第二,教育评价必须遵循一定的价值标准;

第三,教育评价是在收集相关教育资料信息的基础上进行的;

第四,教育评价的手段是科学的;

第五,教育评价的目的是促进学生发展或为教育决策提供依据;

第六,教育评价主要包括学校评价、课程评价、教学评价、教师评价和学生评价等。

为此,本书认为,教育评价是指为提高教育质量、促进教育发展、提供教育决策依据,评价者依据一定的价值标准,在运用多种科学手段收集有关教育信息资料的基础上,对学校整体、课程与教学、教师与学生等方面的变化发展情况进行价值分析与判断的过程。

### (三) 教育评价的功能

#### 1. 调控导向功能

教育评价是一种价值判断过程,因而教育评价具有指挥棒的作用,能够调控并引导教育发展的方向。具体而言,教育评价目标及教育评价指标体系不仅可以为学校指明办学方向,而且可以为教师指明教育教学目标,甚至还可以为学生指明学习目标。

#### 2. 监督检查功能

一方面,教育系统是社会系统的重要组成部分,在教育价值日益凸显的今天,社会各界不仅越来越关心教育的发展,还不断以多种形式的教育评价手段监督检查教育发展的状况;另一方面,教育行政(管理)部门对学校工作进行指导与管理,学校自身对师生的教育教学活

① [美] 格兰朗德. 教学测量与评价. [M] 郑军, 等, 译. 石家庄: 河北教育出版社, 1997.

② 张祥明. 教育评价的理论与实践[M]. 福州: 福建教育出版社, 2000.

③ [美] 布卢姆, 等. 教育评价[M]. 邱渊, 王钢, 等, 译. 上海: 华东师范大学出版社, 1988.

④ 胡中锋. 教育测量与评价[M]. 广州: 广东高等教育出版社, 2006.

⑤ 黄光杨. 教育测量与评价[M]. 上海: 华东师范大学出版社, 2012.

动进行指导与调控,以及教师对学生的进行学习进行监控和指导,都需要借助教育评价。

### 3. 激励促进功能

教育评价的根本目的不是为了评定某种教育现象或教育属性的价值,而是为了促进教育的发展,因而,教育评价能够激励广大被评价者在竞争与合作中不断上进,促进广大被评价者不断提高自身的素质。

### 4. 筛选择优功能

教育评价是一种价值判断,因而,教育评价能够将被评价者分成不同等级。在教育实际中,人们正是根据教育评价的结果对被评价者进行筛选与择优的。

### 5. 诊断改进功能

针对教育评价指标体系,人们不仅很容易发现教育教学过程中存在的问题与不足,而且很容易针对这些问题与不足提出相应的改进对策。

## 三、教育测量与评价的相关概念

### (一) 教育测量与教育评价

教育测量就是对学生们的学习能力、学业成绩、兴趣爱好、思想品德以及教育措施等方面的数量化测定。它主要用于对学生精神特性的测定。

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,通过系统的收集信息资料和分析整理,对教育活动、教育过程和教育结果进行价值判断,为提高教育质量和教育决策提供依据的过程。

#### 1. 两者的联系

教育测量就是指根据一定的理论、规则,运用一定的测量工具对教育现象进行数量化描述的过程。教育评价是对教育现象的状态和价值的判断,这一判断是以对评价对象的客观描述为前提的,没有对评价对象的客观描述,就不会有对评价对象价值的客观判断。

教育测量是教育评价的基础,教育评价要在教育测量所获得的客观信息的基础上进行,同时,教育测量的结果要通过教育评价才能获得实际意义,教育评价是教育测量的具体体现。

#### 2. 两者的区别

(1) 教育测量是对教育现象赋值及获取评价对象数量特征的过程;而教育评价则是对教育现象客体的价值进行判断,是对教育现象的价值关系的认识。

(2) 教育测量是一种纯客观的过程,其突出特点是客观性;而教育评价则具有两种属性,即客观性和主观性。

(3) 教育测量的任务是对教育现象进行量的认识;而教育评价作为一种认识活动,是人的意识对教育实践活动及其结果的综合反映,它反映的是教育活动的过程和终点,又是新的教育实践活动的起点。

### (二) 教育测量与教育测验

测验(test)是用以测量个体的行为或作业的工具。它通常由许多经过适当安排的项目

(问题、任务等)构成,被测试者对这些项目的反应可以记分,分数被用于评估个体的情况。测验通常是教育测验或心理测验的简称,包括智力测验、能力测验、人格测验、成就测验、态度测验、价值测验,等等。各种不同的测验有着各自不同的特性、不同的适用范围、不同的测验规则与程序。

在讨论教育测量与教育测验的关系之前,有必要先介绍一下测验与量表(scale)的关系。量表也称量尺,是确定了测量单位和参照点的测量工具。测验与量表有着密切的联系,但二者存在一定的差别。量表是测验最常用的一种工具,测验则是利用量表等工具进行测量的过程和程序。一项测验可以使用一个或多个量表,也可以不使用量表,而使用图片、实物、器械等工具。量表除了用于测验外,还可用于其他研究方法,如用于问卷调查。

教育测量有广义与狭义之分。狭义的教育测量特指对学生的学习能力、学业成绩、兴趣爱好、思想品德等精神特性的数量化测定;广义的教育测量是指根据测量学的原理,运用测量学的有关方法,对教育现象及其属性进行量化研究的过程。教育测验(educational test)是指对有关个人教育成就情况的衡量,教育测验通常称为教育考试(educational examination)。教育测验既是教育测量的一种工具,又是在教育评价过程中收集相关资料的一种方式。教育测量的工具不仅仅限于教育测验。

### (三) 教育评价与教育评估

教育评价和教育评估是两个常用的概念。有些学者认为这两个概念意义不一样,评价即评定价值,而评估并不表示价值;而有些学者则认为二者没有严格的区别。可见,到目前为止,这些概念的使用还存在着一些混乱。实际上,教育评价与教育评估这两个概念既有联系也有区别。第一,根据《Collins 汉英双解词典》(汉语大词典出版社,1995年版),教育评价和教育评估的英语表示不同,前者通常用“educational evaluation”表示,而后者通常用“educational assessment”表示,既然英语词源不同,那么,其意义则多少有所差别。第二,适用范围不同。教育评价是个比较宽泛的概念,泛指对事物或人物的社会价值做出判断,如学校让学生进行的学业考试,行政管理部门、社会团体对学校的各种评定、评判、评比等。<sup>①</sup>教育评估则是一个比较具体的概念,指对教育领域的某一方面或整体进行一种估量性的价值判断,是一种模糊定量的评价。第三,精确程度不同。教育评价“是按照一定的教育性质、教育目标,采用定量和定性等手段对教育客观事物进行测量、分析、价值判断以及提供决策有用信息的一个过程”<sup>②</sup>,是一种严格而精确的价值判断过程。教育评估则是通过系统地收集资料,采用文字描述的定性分析和信息统计的定量分析相结合的方法,对客体做预测性、估计性的判断的过程,是一种模糊而粗略的价值判断的过程。第四,侧重点不同。教育评价强调完全和精确的感觉,尤其对价值的感觉,这种感觉是靠洞察力和辨别力达到的。评价结果的精确度如何,与评价主体的专业化水平直接相关。教育评估侧重理性判断,尤其按评估标准进行理性判断,这种理性判断是靠理性认识能力才能达到的。评估结果的信度与效度如何,与评估主体的理性认识能力有很大关系。第五,使用情况不同。教育评价的判断标准或原则往往是在新定的或重要的情况下使用,不应根据既定的标准进行评价,而应根据当前的具

① 荀振芳. 大学教学评价的价值反思[M]. 青岛:中国海洋大学出版社,2006.

② 姜风华. 现代教育评价:理论·技术·实践[M]. 韶关:广东人民出版社,2003.

体情况进行评价。如果情况发生了变化,评价标准也应及时发生变化。教育评估则是按照既定的标准进行评估,尽管情况发生变化,但由于标准是事先制定的,因此在某一轮评估中不可能发生改变;如果要改变,也只有下一轮评估中才会发生改变。第六,涉及因素不同。由于教育评价强调精确性,因而,其涉及的因素必然有限,如果涉及因素太多,不仅难以操作,而且也影响评价结果的准确性。教育评估则不同,由于它强调模糊性,因而可以涉及的因素众多,程序也更复杂,教育评价难以判断的问题有可能可以进行教育评估。

尽管教育评价与教育评估有许多不同之处,但它们是紧密联系在一起,通常情况下可以互换。二者的联系机制在于:首先,它们共同作用于一种价值判断,在满足价值主体的需要方面有着共同的目标;其次,没有教育评价,教育评估就会变得狭隘、僵化,最后完全与主体的价值判断、公众的价值追求脱节。<sup>①</sup>

#### (四) 心理测量与教育测量

心理测量和教育测量虽然分属于不同的学科,但是在理论发展与实际应用方面却又是密切相关的。不少人常常将心理测量和教育测量相提并论,认为教育测量就是心理测量在教育领域的应用,其实,两者之间既有联系,又有区别。

##### 1. 两者的区别

(1) 所属学科不同。心理测量是通过观察人的行为,对人的心理特点做出推论和数量化分析的一种科学手段,所属学科是心理学;教育测量是应用各种测试方法和手段,对教育现状、教学效果、学生学业成绩、个人能力、品德等方面进行测定的一种科学手段,所属学科是教育学。

(2) 所关注的侧重点不同。心理测量侧重于考察受试者的能力、人格、态度、兴趣、价值观等个体心理特质;教育测量则侧重于考察受试者对特定知识和技能的掌握,关注教育对个体行为产生的影响及其程度。

(3) 所具有的功能不同。心理测量主要用于儿童智力发育诊断、人才选拔、儿童或成人智力评估、医疗临床诊断等;教育测量主要用来诊断学习行为、评价教学效果、预测学习能力等。

(4) 所用的测量工具不同。心理测量常使用智力测验量表、人格测验量表等量表,或运用调查问卷等测量工具,来完成诊断、预测个体心理特质;而教育测量常使用各种成绩测验,以及各种专业测量机构研发的各种标准化测验等测量工具,用来了解、诊断、评估、预测学习者(或教学者)的学习效果(或教学效果)。

(5) 发展趋势不同。心理测量将继续延续关注个体心理特质的研究路线,随着认知诊断测验理论和技术的不断发展,将更加关注个体内部的认知过程,从而对个体内心的发展变化、判断和预测给出更加细致、更加合理的解释。而教育测量将从单一运用标准化测验、过于重视测验分数的阶段逐渐发展到建立科学、合理、全面的教育评价体系阶段。

##### 2. 两者的联系

(1) 教育测量的理论体系来源于心理测量。现代测量理论和技术的建立,来自美国心

<sup>①</sup> 张继平. 从冲突走向和谐:高等教育评估价值取向的社会学分析[D]. 武汉:华中师范大学博士学位论文,2011:



理测量学家的贡献。1904年,美国心理学家桑代克、教育测量学家麦柯尔先后提出“凡存在的东西都有数量”,“凡有数量的东西都可测量”,这两句话成为教育测量学赖以存在并逐步发展的理论基础。教育测量在心理测量技术发展的影响下,走上了科学化的道路。

(2) 教育测量的技术手段、研究方法来源于心理测量。目前,教育测量研究中普遍使用的各种统计方法、很多统计软件最早也都是来自心理测量学。

(3) 心理测量的最新研究趋势依然指引着教育测量的发展。目前,心理测量正在向认知诊断测验阶段发展,认知诊断测验在教育测量过程中诊断个体学习行为、评估教学效果方面有着其他测验不可替代的优势,尤其适用于目前教育领域里的“因材施教”的教育理念,从而能够从测验这个环节上进一步加强“以学生为本”的教育方针的贯彻与落实。教育测量的目的在于了解学生的发展变化及其成长过程中的各种影响因素,为教育评价和教育决策提供依据,而认知诊断理论和方法的发展能够更好地实现这一目的。认知诊断测验与以往的常规测验最大的不同就是它不再满足于仅仅给应试者一个笼统评价,而是更加关注学习者个体内部的认知过程、加工技能、认知策略、认知模式等,希望通过更加细致的诊断对其认知行为进行矫正、修补、干预等行为,使之不断完善。

(4) 教育测量对心理测量有着重大的影响。从项目反应理论(item response theory,简称IRT,也称潜在特质理论或潜在特质模型,是一种现代心理测量理论,其意义在于可以指导项目筛选和测验编制)的发展轨迹就能看出这一点。IRT理论发端于20世纪50年代,它同时被丹麦统计学家格奥尔拉希(Georg Rasch)和美国统计学家弗雷德里克·洛德(Fred-eric M. Lord)在各自的国家发展起来。尽管采取的研究方法不同,但他们的结果却非常相似。弗雷德里克·洛德在1951年从普林斯顿大学毕业时的博士论文《A Theory of Test Scores》被认为是IRT的开端之作。在随后的30年中,他进入ETS(Educational Test Service,美国教育考试服务中心)工作,并不断深入研究这个问题,其于1980年出版的《Applications of Item Response Theory to Practical Testing Problems》标志着他正式完善了整个IRT理论的框架。因此,弗雷德里克·洛德也被称为“现代测量之父”。

(5) 教育测量和心理测量都关注学习者的学习能力。狭义的教育测量关注学习者的学习能力,而心理测量把学习能力当作一种心理特质。

## 第二节 教育测量与评价的发展历程

历史是一面镜子,可以给人借鉴和引导;历史是过去的沉积,未来的导向;历史是最好的老师,教人融会贯通、惩前毖后;历史是学习的源泉,进步的信心。不懂历史,就难以从中汲取营养,获得经验和启示。了解教育测量与评价的发展历史进程,有助于我们“以史为鉴,明正现实”,有助于我们有效地学习教育测量与评价理论,并有效地开展教育测量与评价活动。

### 一、教育测量的发展历程

对教育测量来说,其发展历程可分为以下三个阶段。