

教育测量与评价学

摇摇摇摇摇摇摇摇主摇摇编摇摇王德清摇摇欧本谷

摇摇摇摇摇摇摇摇副摇摇主摇摇编摇摇曹复恩

摇摇摇摇摇摇摇摇顾摇摇问摇摇乐安中

西南师范大学出版社

目 录

第一章 绪论	(员)
第一节 教育测量与评价的含义	(员)
第二节 教育测量与评价的类型	(猿)
第三节 教育测量与评价的历史发展	(远)
一、中国教育测量与评价制度的起源及其发展状况	(苑)
二、西方教育测量与评价制度的历史发展	(员)
第四节 现代教育测量与评价的特点	(员)
一、现代教育测量发展的特点	(员)
二、现代教育评价发展的特点	(员)

上 篇

第二章 测量的编制	(员)
第一节 测量编制的一般程序	(员)
一、确定测量目的	(员)
二、分析测验目标	(圆)
三、测验题目的编制	(圆)
四、试题的筛选与合成	(圆)
五、测验的标准化处理	(圆)
六、测验的鉴定和使用说明	(圆)
第二节 测验试题的编制技术	(猿)
一、命题的一般原则	(猿)
二、各类试题的编制技术	(猿)

三、论文式测验的编制	(猿)
第三章 摇信度与效度	(源)
第一节 摇信度	(源)
一、信度概述	(源)
二、信度的估计	(源)
三、影响信度的因素	(缘)
第二节 摇效度	(缘)
一、效度概述	(缘)
二、效度的类型及其估计	(缘)
三、影响效度的因素	(远)
第四章 摇项目分析	(远)
第一节 摇难度	(远)
一、难度的意义	(远)
二、难度的计算	(远)
三、难度的转换	(苑)
第二节 摇区分度	(苑)
一、区分度的意义	(苑)
二、区分度的计算	(苑)
第五章 摇学业成绩测验	(愿)
第一节 摇学生成绩测验的性质与功能	(愿)
一、学生成绩测验的性质	(愿)
二、学业成绩测验的功能	(愿)
三、学业成绩测验中值得注意的问题	(愿)
第二节 摇学业成绩测验的类型	(怨)
一、学业成绩测验的分类概述	(怨)
二、语文成绩测验	(怨)
三、数学成绩测验	(怨)
第三节 摇学业成绩测验结果的解释	(怨)

一、年级常模	(怨苑)
二、教育年龄	(怨苑)
三、教育商数	(怨愿)
四、成就商数	(怨愿)
第六章 摇人格测量	(员园员)
第一节 摇人格测量和意义	(员园员)
一、人格测量的含义	(员园员)
二、人格测量的功能	(员园圆)
三、人格测量的分类及方法概述	(员园猿)
第二节 摇自陈量表法	(员园源)
一、什么是自陈量表法	(员园源)
二、自陈量表的编制方法	(员园缘)
三、自陈量表举例	(员园苑)
第三节 摇投射法	(员园远)
一、投射测量的原理	(员园远)
二、罗夏墨迹测验	(员园远)
三、主体统觉测验	(员园愿)
四、其他投射测验	(员园愿)
五、投射测量法的评价	(员园怨)
第四节 摇其他人格测量方法	(员园怨)
一、情境测验法	(员园怨)
二、社会测量法	(员一〇)
三、态度量表法	(员一〇)

下 摇 篇

第七章 摇教育评价的原则与步骤	(员一四)
第一节 摇教育评价的基本原则	(员一四)

一、方向性原则	(页四)
二、科学性原则	(页四)
三、可行性原则	(页四)
四、民主性原则	(页四)
五、激励性原则	(页四)
第二节 瑶教育评价的实施步骤	(页四)
一、教育评价的准备	(页四)
二、教育评价的实施	(页四)
三、教育评价的总结	(页四)
第八章 瑶教育评价的指标体系	(页四)
第一节 瑶教育评价指标体系概述	(页四)
一、教育评价指标体系的概念	(页四)
二、教育评价指标体系的结构形式	(页四)
三、教育评价指标体系的特点	(页四)
四、教育评价指标体系的优缺点	(页四)
第二节 瑶设计教育评价指标体系的依据和原则 ...	(页四)
一、设计教育评价指标体系的依据	(页四)
二、设计教育评价指标体系的原则	(页四)
第三节 瑶教育评价指标体系的设计步骤和方法 ...	(页四)
一、确定评价指标	(页四)
二、分配指标权重	(页四)
三、编制评价标准	(页四)
第九章 瑶教育评价信息的采集与评价结果的计量 ...	(页四)
第一节 瑶教育评价信息的采集与整理	(页四)
一、教育评价信息的种类和来源	(页四)
二、教育评价信息的采集方法	(页四)
三、教育评价信息资料的整理	(页四)
第二节 瑶教育评价结果的计量方法	(页四)

一、累加求和法	(页)
二、加权求和法	(页)
三、模糊综合评判法	(页)
四、标准分数法	(页)
第十章 摇教育评价的心理与调控	(页)
第一节 摇教育评价心理与调控概述	(页)
一、教育评价心理的含义和分类	(页)
二、教育评价心理调控的意义	(页)
第二节 摇主评者的评价心理与调控	(页)
一、主评者团体内部人际影响所发生的评价心理行为···	(页)
二、主评者认知被评对象时所发生的评价心理行为····	(页)
三、主评者评价中自主产生的评价心理行为	(页)
四、主评者评价心理偏差的调控措施	(页)
第三节 摇被评者的评价心理与调控	(页)
一、自我评价的心理行为	(页)
二、受评过程的心理行为	(页)
三、评价结果反馈的心理行为	(页)
四、被评者心理偏差的调控措施	(页)
第十一章 摇学校德育工作评价	(页)
第一节 摇学校德育工作评价概述	(页)
一、学校德育工作评价的意义	(页)
二、学校德育工作评价的范围	(页)
三、学校德育工作评价的特点	(页)
第二节 摇学校德育工作状况的评价	(页)
一、学校德育工作条件的评价	(页)
二、学校德育工作途径的评价	(页)

三、学校德育工作方法的评价	(圆猿)
第三节摇学校德育工作效果的评价	(圆猿)
一、学校德育工作效果评价的内容	(圆猿)
二、学校德育工作效果评价的方法	(圆猿)
第十二章摇学校教学工作评价	(圆猿)
第一节摇学校教学工作评价概述	(圆猿)
一、学校教学工作评价的意义	(圆猿)
二、学校教学工作评价的范围	(圆猿)
三、学校教学工作评价的特点	(圆猿)
第二节摇教学工作的常规评价	(圆猿)
一、教学工作常规评价的意义	(圆猿)
二、教学工作常规评价的内容	(圆猿)
三、教学工作常规评价的实施	(圆猿)
第三节摇课堂教学评价	(圆猿)
一、课堂教学评价的意义和内容	(圆猿)
二、课堂教学评价的实施	(圆猿)
第十三章摇学校体育卫生工作评价	(圆远)
第一节摇学校体育卫生工作评价概述	(圆远)
一、学校体育卫生工作评价的意义与功能	(圆远)
二、学校体育卫生工作评价的标准	(圆远)
三、学校体育卫生工作评价的方法	(圆远)
第二节摇学校体育卫生工作中几个主要方面的评价	(圆远)
一、学校领导体育卫生工作管理的评价	(圆远)
二、中小学体育教学的评价	(圆远)
三、中小学校体育卫生目标管理效能评价	(圆远)
第三节摇怎样参加和开展学校体育卫生评价工作	(圆远)

一、制定自评方案	(猿愿)
二、自评工作过程	(猿园)
三、自评报告的形成及上报	(猿圆)
四、迎接评估组检查	(猿猿)
五、优化整改方案	(猿猿)
第十四章 摇学生综合素质评价	(猿缘)
第一节 摇学生综合素质评价概述	(猿缘)
一、学生综合素质评价的意义	(猿缘)
二、学生综合素质评价的范围	(猿苑)
三、学生综合素质评价的特点	(猿苑)
第二节 摇学生综合素质评价的内容	(猿怨)
一、学生品德评价的内容	(猿怨)
二、学生智力评价的内容	(猿圆)
三、学生体质评价的内容	(猿愿)
四、学生审美素质评价的内容	(猿原)
五、学生劳动技术素质评价的内容	(猿匹)
第三节 摇学生综合素质评价的实施	(猿愿)
一、学生综合素质评价的组织与实施	(猿愿)
二、学生综合素质评价中应注意的问题	(猿园)
三、学生综合素质评价结果的应用	(猿员)
第十五章 摇中等职业技术教育评价	(猿猿)
第一节 摇中等职业技术教育评价概述	(猿猿)
一、中等职业技术教育评价的意义	(猿猿)
二、中等职业技术教育评价的对象和特点	(猿苑)
第二节 摇中等职业技术学校办学水平评价	(猿员)
一、中等职业技术学校办学条件评价	(猿员)
二、中等职业技术学校管理水平评价	(猿原)

三、中等职业技术学校教育质量和办学效益评价·····	(猿苑)
·····	(猿苑)
第三节摇学生专业技术等级评价·····	(猿愿)
一、中等职业技术学校学生技术等级评价的目的·····	(猿愿)
·····	(猿愿)
二、专业技术等级评价的对象·····	(猿园)
三、专业技术等级评价的内容和方法·····	(猿园)
第十六章摇地方政府教育工作评价·····	(猿猿)
第一节摇地方政府教育工作评价概述·····	(猿猿)
一、地方政府教育工作评价的意义·····	(猿猿)
二、地方政府教育工作评价的目的·····	(猿源)
第二节摇地方政府教育工作评价的特点·····	(猿源)
一、宏观管住、微观放活·····	(猿源)
二、外部为主、兼顾内部·····	(猿员)
三、全面评价、突出重点·····	(猿圆)
四、统一要求、鼓励创新·····	(猿猿)
五、重视进展、考虑条件·····	(猿源)
第三节摇地方政府教育工作评价的内容·····	(猿缘)
一、区县政府教育评价工作的内容·····	(猿缘)
二、乡镇政府教育评价工作的内容·····	(猿愿)
第四节摇地方政府教育工作评价的实施·····	(猿园)
一、赋予职权 加强领导·····	(猿园)
二、评管结合 认真自评·····	(猿员)
三、组织力量 严格评价·····	(猿圆)
四、提出意见 改进工作·····	(猿猿)
五、反馈情况 提出建议·····	(猿猿)
六、考核奖惩 增强实效·····	(猿源)
后记·····	(猿愿)

第一章 绪论

第一节 教育测量与评价的含义

所谓“测量”（~~测量~~）本来是自然科学的方法，就是按一定的测量尺度获取数量结果的步骤。不过，智力测验和学力测验一类的教育测量的尺度，不能像测量重量、长度之类的物理测量的尺度，诸如米尺和公斤秤那样，足以供精密而普遍地进行测量使用。例如：在教育测量中的成就分数，五阶段法（标准评分）、百分位数等等，就是属于这种尺度。这些都是依据一定的常模组（~~燥电早燥~~）的实验，任意地制造的尺度，它的应用范围有限。而且这种尺度的单位，不像物理尺度中诸如厘米、米，克、千克那样绝对、客观。从~~员因~~年第一个作为测量人的智力工具比纳——西蒙智力量表问世至今，各种测验面临着两种对它的应用和发展有害的态度：一种是未思索其原理、内容、方法就妄加批评；另一种是盲目迷信，夸大测验功能，且对测验结果极不慎重。这两种态度的共同原因之一都是对心理测验的性质缺乏足够的认识。测量在教育中的运用，我们一般称之为教育测量，旨在测定学生对某学科或某组学科经过学习或之后所获得的知识 and 技能，从而为评价教育目标实现的程度提供依据。教育测量可以按照大、中、小学所设置的课程编制各学科的测验，也可以对一组学科编制综合性的测验，又可以为了发现学生的掌握知识或技能上有那些缺陷而编制诊断测

验。凡此种种测验,都应像智力测验一样,在测验的编制、实验、记分、解释上控制误差,即要遵循标准化原则。

教育评价在本世纪 70 年代开始受到世界各国的广泛重视,到了 80 年代,即已成为一门具有独立研究价值的教育科研领域。但由于它是一门新兴的学科,在理论和实践上仍处于积极摸索和有待完善的阶段,因而至今还没有一个确切的、严谨的、被一致接受的科学定义。人们通常认为,评价意指判断一些事情的价值,在教育中的评价,可以说就是对照教育目标对由于教育行为而产生的变化进行的价值判断。具体来说,教育评价是人们按照一定社会的教育性质、教育方针和教育政策所确立的教育目标,对所实施的各种教育活动的效果及学生学习质量和发展水平进行科学的判断。判定的目的在于使评价者对教育活动效果进行估量,探索教育规律,从而更有效地指导教育活动,使被评价者增强自身辨别力,调正自己的行为,以符合既定教育目标的要求。^①

教育评价与历来的考查办法在背景、范围和对象等方面均有明显的不同。考试只能考查记忆、理解多少,却不能考查态度、兴趣、品德、性格、情操等方面的效果。根据现代教育理论来看,考试的范围已经过于狭窄。教育评价则是以教育的价值观点为标准,来评量教育效果,它要求对教育和学习作全面性考查,其范围包括:学习方面的知识、理解、技能和认识能力;情意方面的态度、兴趣、习惯、鉴赏;个性方面的性格、道德品质等。当然教育评价并不能取代其它考查方法,而应是根据评价对象的需要,选择适当的方法。如考查学习效果就需要用考试法,考查兴趣、态度等就需用观察法、测验法、调查法等等。

在教育评价的发展史上,“测量”的概念率先问世,后来才强

^① 傅亚宾、张玉田:《略论普通教育评价》,《东北师大学报(教育版)》,1983年第 7 期。

调了“评价”的概念。因此,揭示这两个概念的异同与关系,对于明确教育评价的意义极为重要。可以说,评价与测量都使用测验,比较系统地收集学生的作业等资料分析,推断出结论。但有测量不一定有评价,为了评价又常常需要测量。测量的基本要求是科学性、客观性、准确性,通常以效度、信度、区分度做为衡量测量的指标。教育上的测量是客观地、正确地将学生达到教育目标的程度加以数量化。评价需要基准,评价的基准是价值,教育评价则是以测量结果有多大的教育价值为其课题的。因此,只有测量而无评价,不仅不能说明问题,而且滥用测量还会产生许多不良影响。例如,衡量一个地区、一所学校的教学质量,虽然测试工具、实施程序等完全符合科学要求,统计检验的结果证明各区、各校的学生水平确实存在显著差异。这能否解释为某区、某校的教学质量高或低呢?从评价的角度看,这些仅仅是一些数据,评价的任务在于解释这些数据,给予价值判断。^①另外,评价与研究的主要区别在于:研究的结论往往要求具有较大的概括性,不一定能直接用于实践,而评价的结论则是特定的,直接为教育决策服务。也可以说,教育评价是运用一切可行的评价技术去评量教育所产生的预期的一切效果。

第二节 教育测量与评价的类型

教育测量的类型,是根据不同的分类标准而有所不同的。在教育过程中常见的是以测量的对象来加以分类,教育测量有:

(一) 学业成绩测验 这类测验是测量学生的知识、技能。

^① 张天麟《教育评价研究中的几个问题》,1984年湖北襄樊“普通教育质量评价研讨会”论文。

(圆) 智力测验 这类测验用于测量学生的观察、识记、分析、判断、推理等思维活动能力。

(猿) 人格测验 这类测验是用于测量学生智力以外的心理的各种特性。

(源) 特殊能力测验 这类测验是用于测量学生某种特殊能力的程度。

教育测量作为改进教学的良好工具,对提高教育质量起着巨大的作用,同样教育测量也是进行教育科学研究不可缺少的科学方法。当然在运用这种工具的过程之中,我们应采取严肃的态度,充分认识到教育测量还有许多尚需完善的地方。就世界教育测量的现状来说,关于知识和技能方面的测量,已经有了一些比较有效的测量工具,但对学生的智力和人格方面的测量发展得还比较迟缓,测量的结果精确性也较差。任何一种工具,只有在使用中才能发现它的缺点,并加以不断改进和完善,教育测量也是如此。

当前,不同的学者在从事教育评价研究中,从不同的角度和层次,提出了不同的教育评价的分类或提法。布卢姆根据教学活动的要求,把评价分为三种不同的类型,即形成性评价(云)、诊断性评价(粤)和总结性评价。另外,有些学者根据评价的对象,把评价分为广义教育评价和狭义教育评价。广义的教育评价是以教育的全领域为对象,它涉及到教育的一切方面,教育与社会、教育与经济、教育目标、课程设置、教学内容、教学方法、品德教育、教育管理、学校结构、教师活动、师生关系以及社会影响等各方面。狭义的教育评价是以学生为对象,专指学生的学习和生活领域中,对学生的智能发展、个性形成、身体发育等方面的评价。根据评价的基准来划分,又可分为相对评价(又称相互参照评价)和绝对评价(又称客观标准评价)。相对评价是指在某一集团中(班级、学校、地区或国家),以这个集团的平均状况为基准,评价每一个学生在这个集团中所处的相对位置。

绝对评价是以预先设定的学习目标(从教师一方说,即教育目标)为评价基准,评价每一个学生的到达程度。就目前而言,教学中有以下四种常见的评价类型:^①

(员) 测定性评价,是在教学前进行的,其中一种是关心学生对教学的准备状态,即是否具有必要的预备性知识和技能,这种掌握性的测验试题难度较低,常常属于标准参照性的测量。另一种则是关心教学计划对班级的适合程度,确定学生原有的实际水平,这种检查性测验试题总难度属中等,常常属于常模参照性测量,以便了解学生的实际水平和差异。

(圆) 形成性评价,是在教学各阶段中进行的。目的是了解学生的学习进程,向师生提供反馈,以便改进和指导教学,不注重对学生进行分等。采用掌握性测验,试题是根据学习任务的要求,可易可难,常属标准参照的测量。

(猿) 诊断性评价,在需要时(大多在形成评价后)即可进行。其目的在于确定学生最常见的学习错误,以便进行补偿性的教学。测验题是为鉴别学生学习错误而特别设计的,其内容有限、难度较低。

(源) 终结性评价,在课程与单元结束时进行的。目的在于给学生作出鉴定或分等,了解教学目标达到的程度,评价教学是否有效。学校中的期中或期末考试常属于此类,采用常模参照性的检查性测验,涉及面广、试题难度不等。

教育评价是按照一定的教育目标,运用系统的科学方法对教育现象进行价值判断,从而使教育工作不断得到改进和提高的过程。当前,建立系统的教育评价制度,不仅是我们端正教育思想、加强教育管理、提高教育质量的有效手段,而且能够为普及九年制

^① 王钢《评价的一般方法与步骤》,《考试研究》,1982年11月。

义务教育,按照“三个面向”的方针深入教育改革,保证我国教育沿着社会主义方向前进提供科学的依据。根据各国实施教育评价看,一般来说教育评价具有以下功能^①:第一,通过教育评价可以诊断学生学习的进步、缺漏或迟滞情况,为有效的指导和治疗提供客观资料;第二,通过教育评价可以了解学生的适应状况、态度、兴趣、价值观、智能、性格、学业、成绩、身体与环境条件等方面的情况,以便因材施教;第三,通过教育评价可以诊断课程和教材、教法的优劣,以便于及时改进;第四,通过教育评价可以帮助学生确认自己的个性、能力、兴趣,激发学生的学习动机;第五,通过教育评价使各学校的教育活动都能以全面发展受教育的才能为目标,用一套新的方法,全面评定教育目标是否实现。

第三节 教育测量与评价的历史发展^②

近半个世纪以来,教育测量与评价无论从理论上还是从实践上,无论在测量与评价目标的确定还是在手段的选择上,都有了很大的发展,并且被越来越多的国家所接受。教育测量与评价理论与制度是多年来教育科学工作者们经验与教训的总结,对教育测量与评价制度发展历史的回顾,对尽快建立符合中国特色的教育评价制度有重要的启示意义。

① 李聪明《教育评价的理论和方法》,台湾幼狮书店,1980年版,摘引其主要内容。

② 第三节的一、二两个问题曾以论文的形式发表在《山东师大学报(社科版)》1980年第 3 期上,题目为《中外教育评估制度发展史简述》,此处略有修改。

一、中国教育测量与评价制度的起源及其发展状况

(一) 教育评价源于中国古代的科举制度

摇摇美国一些知名教育评价学者杜波依斯(孕·匀·孕建译)、波普纳(宰·允孕建译)、活森(月·砸·宰燧译)等人研究,认为教育评价是起源于中国古代科举制度,这种制度是用一套完整的考试制度和方法来挑选人才,含有教育评价活动的萌芽。考试是一种古老的制度,统治阶级常用考试的办法为其选拔统治人才。我国是最早采用考试选拔人才的国家之一。在古代的学校中建立了定期对学生进行考试的制度,并在实践中创造了许多考试方法。

据《礼记·学记》记载:我国古代的西周时“国学”中的大学,就设有定期的学业考查制度。汉朝的大学生也只有通过考试才可以毕业授职。在西汉中期规定为每年考试一次,考试的方法有:“口试”、“策试”和“设科射策”三种。元朝的国子学把考试分为私试和升斋考试。私试每月举行一次,试卷上等的给一分,中等的给半分,一年内积分到八分以上升为高等生员;升斋考试每季举行一次,下斋生考试成绩优等的升中斋;中斋生考试成绩优等的升上斋。到了明清后,我国古代的学校教育逐步演变为科举制度的附庸。与此同时各类学校的考试制度,在原来的基础上和科举制度的影响下,在考试内容和方法上也有了发展。从隋唐至清末科举制度不断完善,持续达一千三百多年之久。形成了一系列有关教育、学力检查和评价制度,具有一套等级划分、考核标准、实施办法和选择目标体系。早于中世纪西欧各国的就业资格考试的科举制度,是以封建主义的价值观为准绳,把政治审查、学力测评与健康要求结合在一起的一套独立的评估制度。科举制度中,考试的主要内容是儒家经典,使用的方法不外乎口试、帖经、墨义、策问、诗赋等。科举制度的基本特征是通过逐级的考试选拔人才。到贞观元年,袁世凯奏请废止,作为一种最初的萌芽,它在教育评价史