

## 绪 论

### 第一节 教育评价的历史沿革

教育评价是从教育测量中发展起来的。教育测量最早可以追溯到 19 世纪的各种测验，而对学校或其他教育机构最初的成就评价，可以追溯到 19 世纪上半叶，距今已有一百四五十年的历史。它的起源与发展和对学生的学力检测活动的发展有着紧密的联系。

#### 一、教育评价的萌芽时期

教育评价源远流长，从学生成就测验开始，至今已有几千年的历史。迄今为止，学术界对其源头的产生有不完全一致的看法。共有两派的观点：

一派是以日本、中国学者为代表，认为教育评价最初萌芽源于中国古代的考试制度，它被公认为世界上最早的一种完备的考试，有人称其为“文官考试”。但是，由于种种原因，20 世纪以来，我国对教育测验、测量、评价的研究一直未能发展起来。

另一派是以西方学者为代表，认为教育评价萌芽最初源于西方，1702 年英国剑桥大学首先以笔试代替口试进行的入学考试，被喻为西方考试的先河。1845 年美国的波斯顿州的笔试入学考试，特别是 19 世纪后半叶以来，又使得美国教育评价得到迅速的发展。

在 19 世纪上半叶和更早的时期，美国的学校普遍感到选拔学生的考试缺乏标准，因而也缺乏考试应有的客观性，其基本的方法是对学生逐个进行口试。以后，由于学生人数的日益增多，这种形式的考试实际上已经很难实施了。此外，这种考试并非是标准化的，故而给成绩的评定也带来了很大困难。

1845 年，在美国著名教育家贺拉斯·曼（Horace Mann）的倡导下，美国波士顿文法学校首先引入了书面考试，这就使以统一的试卷测试众多的考生成为可能。此后，书面考试对西方的测验方法有了很大的推进，测验的手段更加灵活，但是成绩评定的客观标准问题仍然未能很好地解决。针对这种情况，1864 年，英国格林威治医学院教师乔治·费舍（George Fisher）收集了许多学生的成绩，汇编成了成绩量表，叫做量表集，试图为当时的考试提供一个可供参考的标准。费舍先生的工作事实上是用科学方法研究教育测量问题的最初尝试。但是，由于种种原因，费舍先生的这一工作并没有引起教育界的更多注意，故而没有更多的人继续他的研究。然而，在教育界引起人们对考试问题极大关注的是美国著名的莱斯（Joseph Rice）拼字测验。

在 1897 年，莱斯发表了他对 20 个学校的 16000 名学生所作的拼字测验的结果，测验表明：8 年中每天用 45 分钟的学生与每天用 15 分钟进行拼字练习的学生的测验成绩并没有多大区别。莱斯的这一结论在教育界引起了强烈的反响，引起了人们对测验方法的普遍关心。

除此之外，教育测量在这一时期之所以引起人们普遍的兴趣，还有另外一个方面原因，心理学尤其是实验心理与心理测验的发展对此起了极大的推动作用。在德国，1879 年冯特（W. Wundt）在莱比锡大学首创了心理实验室。在实验心理学的研究中，实验心理学家逐步摸索出了一套测量的方法，这套方法对教育测量的发展有很大的影响。在英国，19 世纪下半叶的心

理学家对个体差异的研究有很浓厚的兴趣，其中对教育测量的发展有较大影响的是 1869 年高尔顿 (F. Galton) 的研究结果的发表。他运用统计学的方法提出了一个由他构想出来的量表。他认为，人的能力分配是常态的，大部分人的能力是正常的，而一少部分人的能力是超常和低常的。高尔顿的这一思想对当时的心理学与教育学界都颇有影响。早年留学德国的美国著名心理学家卡特尔 (J. Mck. Cattell) 受其影响也很大，卡特尔指出：“心理学若不立根基于实验的测量之上，决不能够有自然科学之准确。”所以他十分注意对心理测验的研究。1894 年后，卡特尔在哥伦比亚大学执教时，他就试用各种测验考核本校一年级与四年级大学生。在 19 世纪的最后十年，各种心理测验随着心理实验的发展受到了人们越来越多的重视。

## 二、教育测量时期

从 19 世纪末到 20 世纪 30 年代教育评估进入一个新阶段，这一阶段前后约 50 年时间，这一阶段称为测量运动阶段。学生个体学力测验的客观化、标准化问题是这一阶段的中心问题。这个时期的主要代表人物是桑代克 (E. L. Thorndike)，他于 1904 年发表的《心理与社会测量导论》一书成为这一时期的标志。在这本书中，桑代克系统地介绍了统计方法以及编制测验的基本原理，并提出了一个著名的论断：“凡是存在的东西都有数量，凡是有数量的东西都可测量。”这一论断对教育测量的发展起了很大的推动作用。

教育测量这一阶段的客观化、标准化问题受到了极大的重视。所谓客观化表现在两个方面：

第一，标准客观化。每一测验都在较大的范围内进行实测，从而确定集体常模，由此可以获得客观等距性的标准量表。

第二，评分客观化。这种测验的突出特点是题目的回答不要

求考生长篇大论的书面文字，而是要求明确、简单回答，评分客观、准确的所谓客观式测验。<sup>①</sup>

在这一时期人们的兴趣主要集中在各种标准化测验编制的研究上。

20 世纪以来，科学技术的发展越来越成为生产发展的关键，使得人才的需求量不断增加。为了更好地鉴别和选拔人才，出现了改革传统考试制度的趋势，产生了智力测验和教育测验的浪潮。教育测验来源于心理测验，是对人的智力和能力的测验。教育测量时期从 1916 年至 1930 年，约 15 年时间。1916 年在心理学界，斯坦福量表的问世，标志着心理测验的基本定型。在对身体有缺陷儿童的研究中，法国心理学家比奈（A. Binet）1906 年提出了他的第一个智力量表，这个量表已经指出 3 岁、5 岁、7 岁以及 9 岁的儿童在量表上可以做到什么。此后比奈于 1908 年和 1911 年两次修改了他的智力量表，其中 1908 年的量表引入了智力年龄的概念。1916 年美国斯坦福大学教授推孟（L. M. Terman）主持修订了比奈量表，推孟在智力量表里首次引用了“智商概念”，从而使心理测验达到了较为成熟的阶段。

心理学与教育是密切相关的，心理学对测验的研究成果很快就被用到了教育测验之中。在心理测验之后，教育测量在十多年中也迅速地发展了起来。在广泛地开展研究的基础上，各种标准化试卷被大量地编制了出来，各种关于教育测验的出版物在这一时期也极为活跃。这一时期是测验研究的全盛时期。对学校以及各种教育机构成就的评价，无论在内容还是在方法上，都受到测验研究的影响。当时将智力测验的方法用于考查学生成绩上叫作成就测验或学力测验。例如，算术测验，是从易到难提出 12 个问题，测 15 分钟，对象是小学六年级学生。此后，这种测验逐

步发展成标准化测验。1923年，出版了斯坦福成绩测验，它可以综合运用到全部课程上，是第一个标准化的成绩测验。这个测验还从有代表性的1500名儿童中得出了各种年龄的平均数，作为国家标准，以便于比较。在此期间，不仅学生的学业检查应用测验，就是教育诊断、课程调查和教育实验，也都开始应用教育测验，连教师们用来检验教学效果的考试，也提倡实行客观测验。

在19世纪初，英国就已经有了评价教育、公共卫生等事业的活动。这种活动通常是由政府指定一些委员会来进行的。但是这一活动基本上是凭个人的主观印象，以非正规的程序进行的。

### 三、教育评价时期

标准化的测验只着眼于客观数字的测定，虽然使考试客观化、标准化了，而且能够把学生的能力换算为数字，测定出学生之间的个别差异，但是测验的种类毕竟有限，不能反映教育的全面效果。例如，学力测验只能测出那些简单的、偏重记忆的、能够用数字表示的部分，而对于诸如社会态度、思想品质、实际技能、创造能力、兴趣、鉴赏力等，就很难变成数字，给予量化。智力测验的分数往往不能反映一个人的比较复杂和高级的认识能力与技能。弗斯特1950年发表的一篇文章《智力测验与批判性思维及知识测验之间的关系》，认为智力测验与比较复杂的认知能力和技能的测验之间相关较低，因此，测验很难测出学力的全部。而且智力测验的结果表明，一个人的智商基本上是恒常的，也就是智力测验的分数在儿童早期（3~7岁）是可以改变的，但以后的学习很少能使智力测验的分数得到改变。这样一来，就很容易使教师把学生看死了，把智力测验分数低的学生看成是不可救药的，把智力测验分数高的学生看成是尖子。但是事实并非如此，人的智力是可以改变的，通过后天的学习可以将人

的智力开发出来，可以发展人的能力。例如，根据布卢姆的调查研究，学生智力测验的分数与以后学生成绩的相关约为 +0.50 至 +0.70。而学生学习某学科所获得的准备条件（例如，学习乘法要有学好加法的准备条件，学习除法要有学好乘法的准备条件），与学生学习某科成绩的相关一般为 +0.70 或更高一些。这就说明，学生学习某科所获得的准备条件对其今后学习的影响，比智力测验的分数更重要。为了解决测验的局限性，许多教育家就开始倡导教育评价。<sup>①</sup>

由于杜威等人倡导实用主义心理学，提出了教育的目的在于生活和以儿童教育为中心的观点。1918年，受杜威影响的教育家在芝加哥开会，组成了进步主义教育联盟。该联盟对偏重于记忆的考试和测验进行了批判。1933年，由进步主义教育联盟的会长艾钦（W.M.Aikin）等主持，开始了一个批判教育测验的“八年研究”（1933~1940年）。于是，1933年就作为“教育评价”一词产生的时间。在研究中，他们认为教育测验有很大的片面性，不能全面反映受教育者各方面的实际学习效果，教师以教科书为中心所采取的考试方法和测验方式使受教育者易于养成被动学习的态度，不符合学生的发展和社会的需要。因而，他们在对课程内容进行改革实验的同时，特别重视提出一套有别于教育测验的成绩考查方法。以泰勒（Ralph.W.Tyler）教授为主的评价委员会，在“八年研究”中，专门从事了这一工作。他们为新课程的设计了一套教育评价方法，经过八年的实验，证明教育评价促进了新课程的实验效果。该委员会最后提出了一份“史密斯-泰勒”报告，被称为是“划时代的教育评价宣言”。这份报告，第一次提出了教育评价的思想。

泰勒的教育评价思想认为，教育不能只是知识的灌输，而要

陈玉琨：《教育评估的理论与技术》，广东高等教育出版社，1987年版。

以全面发展受教育者的才能为目标，所以需要一套新的教育评价方法，对教育效果进行全面判定。在进行教育评价时，并不排除使用一定的测验方法，但必须综合利用各种方法，来全面评价教育目标是否达到和达到的程度，以及原定目标有无进一步修改的必要。泰勒的研究结果认为，教育考查的方法，对于实现一定的教育目标，有着极大的影响。如果使用不恰当的成绩考查方法（如激烈的升学竞争考试），往往会促使教育走向不正确的方向。象单纯地注重知识教育，可能就是它的严重后果之一。他们不仅把教育评价同教育目标紧密联系在一起，而且还设计了一种叫“问题情境测验”的方法，来考查学生的创造思维、逻辑思维、分析和解决实际问题的能力。他们还注重了全面地、一贯地针对德、智、体全方面的要求，进行教育评价；注重了对学生成绩的分析，以便进行有针对性的指导；注重了学生的自我评价和相互评价。这种教育评价，打破了过去考试只注重分数，排名次的简单做法，而注重了学习成绩的全面考查和分析。这种评价，也扩大了范围，既包括了对学生的学习效果、智能、身体健康状况、思想道德水平、志趣和理想的评价，也包括了对学校、教师、课程和教材的评价。

1942年泰勒教授撰写的“八年研究”报告中提出的教育评价思想，首先在美国引起很大的反响。根据他开展的教育评价活动，能够获取大量的信息、能够准确地判断教育目标完成的程度，这有助于全面地分析教学工作，明确改进教学工作的方向，在提高教学质量中起到很好的作用。此后，教育评价的思想与实践，很快为全美国的许多学校所采用。不久，它又相继被传播到西欧一些国家。

第二次世界大战之后，一些国家面临着恢复与发展经济的任务，他们为了社会的发展、经济的繁荣，都热衷于增加教育投资以改变其落后面貌，那么，教育投资的效果究竟如何呢？各国都

希望通过教育评价，来促进提高教育投资的效益、推动教育事业的健康发展和不断提高教育质量。特别是二战后日本全面实施了教育评价，使教育在恢复和发展日本经济的过程中，起了特殊重要的作用。

#### 四、现代教育评价时期

20 世纪 60 年代至今，被称为现代教育评价阶段。教育评价有了进一步的发展。纵观教育评价的发展过程，可以看出其显著特点，这就是由只注重追求客观性到同时也注重追求教育性的过程，是由着重强调鉴定、选拔学生到不断改进教育、创造适合受教育者的教育的过程。<sup>①</sup> 为了认清教育评价发展过程的特点，我们可从下列几个方面进行具体分析。

##### 1. 从单纯教育测验到早期教育评价

在 19 世纪末期，由于德国实验心理学的发展，英国学者对个性差异研究的成果以及法国对有生理、心理缺陷学生的研究，使教育测验有了很大的发展。特别是 20 世纪初的美国，教育测量运动蓬勃发展起来，据统计，美国在 1928 年有 3000 多种测验方法。但是，人们从实践中逐渐认识到，这些测验方法都有较大的局限性，由于它过分地追求对教育现象的数量化描述，而有些属于教育领域，很难直接量化，且对提高教育质量有重大影响的内容，如态度、兴趣、情感、意志等，却被排斥在教育测验之外，加上当时的教育测验方法又多是偏重于考察记忆的内容，这样就使教育评价理论的发展受到制约。自从“八年研究”开始的早期的教育评价直至今日，教育评价不仅重视对教育现象的数量描述，而且更加重视获取与分析信息，对评价结果给予全面的解释和价值判断。比如评价学生某科的学习质量时，不仅要通过测

验的方法获得学生该科的学习成绩，还要重视采用其他可行的技术与手段，对学生应用知识、分析与综合、评价等较高层次的认识能力，以及学生的兴趣、态度、情感、意志等方面的内容进行全面的评价，从而获得该生这门学科学习质量的比较科学的结论，同时，也有利于改进教与学的工作，促进学生进一步提高该学科的学习质量。

### 2. 从相对性的评价过渡到绝对性的评价

从相对性的评价过渡到绝对性的评价是教育评价思想的一个重要发展。所谓相对性评价，是采用相对目标和标准来进行测定与评价。例如评价学生的学习质量时，不是就本人的实际学习水平进行评价，而是用跟其他人的差异（即在正态分布中的不同位置）的比较来表示。显然，这种评价只能反映出学生的学习水平在集体中的名次，而不能表示出学生的实际学习水平，也难以达到促进学生积极学习的目的，这种评价在一定程度上会扼杀教师与学生的积极性。因为一般来说，无论教师如何认真教学，无论学生如何刻苦学习，在一个集体中，学生的学习质量总是有差异的，好学生往往总是名列前茅，而学习差的学生虽然经过努力，学习成绩有了明显的提高，但随着集体水平的提高，他们常常仍处在最后的名次。

为了追求教育评价的教育性，就逐渐采用绝对目标的标准来进行测定与评价，这就是所谓绝对性的评价。这种评价能够比较客观地反映出学生的实际学习水平，达到目标的具体程度，因而就能很好地调动师生的积极性，就能发现教与学的问题，从而促进学生不断提高学习质量。可见，这种评价具有较强的教育性。

### 3. 从拟定一级教育目标到建立教育目标体系

早期教育评价，一般是根据对象拟定若干一级教育指标进行评价，而建立教育目标体系，则使教育评价得到突破性的发展。

1949年在波士顿举行的“美国心理学大会”上，与会者提出对组成教育目标的各种因素进行分类，并逐步建立教育目标分类学的重要任务。与会者认为，只有建立教育目标体系，进行教育研究工作才有共同的基础，才能深入地开展教育评价工作。

在创建教育目标分类学的过程中，布卢姆教授作出了杰出的贡献，1956年布卢姆发表的《教育目标分类学》，奠定了教育评价的重要理论基础。它在促进教育评价理论的发展中起了重要作用。他因取得研究教育目标分类学的成就，而受到美国乃至世界各国学者的赞扬，赢得了学术声誉。布卢姆教授对教育目标的分类方法及其成果，被列为美国教育研究的重大成果。它对于完善教育评价的理论、有效地开展教育评价工作，起了很重要的作用。

#### 4. 从早期的总结性评价到现代的形成性评价

早期的教育评价目的，多是评价教育活动达到教育目标的程度，且一般都是在某项教育活动结束之后进行的，所以把这种评价称之为总结性评价。由于这种评价是在某项教育活动结束之后进行，因此，它不能提供教育过程中的反馈信息，故不能直接起到优化教育过程的作用。

在教育评价实践中，人们逐渐认识到总结性评价的不足之处。根据教育评价的改进与优化教育的功能，1967年斯克里文(M. Scriven)指出，除了总结性评价之外，还应有一种在教育或教学过程中进行的评价，其评价目的主要是提供大量的反馈信息，用以调节、控制、优化教育与教学过程，创造更加适合教育对象的教育，这种在教育或教学活动进行过程中的评价，被称之为形成性评价。

从总结性评价到形成性评价，也是教育评价思想与实践的一个重大发展。形成性评价经过布卢姆运用于教学评价活动之中后，内容不断得到丰富与发展。目前，它已受到普遍重视，成为

教育评价中一种较为广泛使用的评价。

## 第二节 评价及教育评价的含义

### 一、评价的实质

“评价”是一个使用范围相当宽广、几乎人尽皆知的词，如《辞海》所述，它是泛指衡量人物或事物的价值。据此可以说：“没有评价，就没有生活，就没有一切。”衣、食、住、行、学习、工作、婚姻、成就、人际交往、社会等等，都要选择评价。一般来说，人们每时每刻都从外界接收到大量信息，其中一部分信息经过分析、综合等思维过程，形成某种判断，进而进入决策阶段。这种分析、判断、决策过程，必然蕴含着各自具有的不同动机和目的，针对不同的人物和事物，当然就会产生不同层次、具有不同效果的评价过程。

对人物或事物作出主观的“是好是坏”的判断，往往被人们称之为“评价”。这里言及的“好”或“坏”，就是指“价值的判断”。“价值”一词在《辞海》里是这样解释的：指事物的用途和积极作用。对于每个有意识的人来讲，由于经历的不同，经济地位的不同，所获经验的不同，形成的意识、情感、情操也就不同，就会产生不同的习惯、爱好、志趣、利益和欲望，在各种活动领域中存在着程度大不相同的需要和追求。也就在这时，社会上的人们就形成了不同层次的千差万别的价值观。有的侧重考虑功利方面的得失；有的志趣在于对科学方面或艺术方面或道德方面的真、善、美的追求；也有的则倾向于某种信仰上的忘我献身。这种价值观的不同层次，就带来了有时令人快慰、有时令人烦恼的后果。人们对同一个人物、同一件事物，可以作出完全不同的以至截然相反的“价值判断”。在教学评价过程中，这种现象表露得最为突出的科目是“作文”，不同

的评判者，存在着不同的倾向：或着眼于题目的理解，或着眼于词句的练达，或着眼于文章寓意的深刻等。这就涉及到评价过程中的主观性、不一致性，这种主观性、不一致性是绝对的。

我们追求科学的评价，它是在人们的主观认识最大程度地符合客观现实的情况下，作出的价值判断。因而，这种价值判断，是比较客观的，比较一致的。在这里，并不意味着评价者在价值观上取得了完全统一，因为这是根本办不到的。在实际评价工作中，往往在进行评价设计时，适量减少一些在价值标准上有过大争议的评价目标和评价指标，并要详尽地表述评价指标，以期作出评价者一般都可以接受的价值标准方面的约定，从而求得评价标准的大体一致。

换句话说，主观评价反映客观实际的前提，首先应该是在价值观的较多层次方面，大家大体相近，这样才能比较容易地形成统一的且比较正确、比较完备的价值规定。这种评价就是本书所要着重研究的“评价”。不过，从上述可以看到，现实中的“科学评价”已隐藏有不科学的成分，客观的评价是以评价者认可的评价目标和评价指标而定的。看来，对科学的客观评价的弱点，用经验的主观评价来补充，即以某种形式把两者结合起来作为决策的依据，才可能是更客观、更科学的。

## 二、教育评价的实质

在我国文字中，“评价”是评定价值的简称。在英语中，“Evaluate”（评价）这个词，在词源学上的含义也就是引出和阐发价值。从本质上来说，评价是一种价值判断的活动，是对客体满足主体需要程度的判断。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。① 比如，教师通过辛勤的劳动为社会培养了人才，评价教师的工作就是对教师是否对社会作出了贡献、他的贡献有多大作一个价值判断。教师评价的目的是进一步促进教学、教育质量的提高，使教师的工作能产生更大的价值。事实上，在教育评价的各种界定中，美国学者格朗兰德（Gronlund, N.E.）在 1971 年作出了一种极为简洁的表述。评价可以简单地用下式表述：

评价 = 测量（量的记述）或非测量（质的记述） + 价值判断

这就是说，评价是在量或质的记述的基础上进行价值判断的活动。格朗兰德的这一公式虽然粗糙了一点，但它确实抓住了评价活动的本质。

对事物进行量（或质）的记述，我们称为“事实判断”。事实判断是对事物的现状、属性与规律的客观描述。比如，我们说，这个学生已经掌握了四则运算的规则，这就是事实描述，它记述了学生知识掌握的现状。在教育活动中，教育测量与教育统计都属于事实描述的范畴。事实判断的基本要求是它的客观性，即真实地反映事物的本来面目。“价值判断”则不同，它并非是由客体单方面所决定的。如前所述，教育的价值由教育活动满足主体需要的程度来决定。不同的主体由于其需要不同，对教育活动就有可能产生不同的判断。

价值判断是在事实描述的基础上，根据评价者的需要和愿望对客观事物作出评判。由于它受评价者价值观念的制约，因此价值判断有一个显著的特点：它是一种客观性与主体性统一的活动。所谓客观性，就是说它是在客观地描述对象的基础上进行的；所谓主体性，就是说评价的结论又与评价者本身对事物“应

该怎样”的认识有关，反映了评价者的主体需要和愿望。格朗兰德的公式表述的就是这样一个关系。

### 三、教育评价的特点

人们对教育活动开展的普及，教育观念的不断更新，对教育评价的理解越来越深刻，教育评价的理论也得到了发展。现代的教育评价发展呈现如下趋势和特点：

#### 1. 教育评价内容的广泛性

评价内容的广泛性，这是现代教育评价的重要特点。教育评价，就其对象领域而言，有狭义和广义之分，狭义的教育评价范围，一般只以学生为对象；广义的教育评价范围，是以教育的全部领域为对象的，它包括与教育对象相关联的各个方面的实态把握和价值判断。现代教育评价强调其内容范围的广泛性。

早期的教育评价的范围主要局限在学生身上，而现代教育评价虽然重点是放在学生身上，源于学生的质量是教育质量的根本标志，这是由教育的性质决定的。但是，评价范围又不局限于学生，而是把教育的全部领域作为评价对象，即包括了促进学生发展变化的各种因素。这是因为学生的成长与发展，是受多种因素制约的，除了学生自身的因素以外，还有其他方面的许多因素，诸如人力、物力、财力，等等。简言之，学生质量，是其主观因素与客观因素、学校因素与家庭及社会等各个方面的因素共同作用的结果。

从评价的具体内容上看，传统教育评价偏重于智育，强调考评学生记忆教科书的程度，教师怕学生把知识遗忘了，便进行频繁的考试，由此加重了学生的紧张心理，也导致了师生的情绪对立。现代教育评价着眼于人的全面发展，从德、智、体等多方面提高学生的素质。教师为了综合地评价学生的质量，可以采取多种评价手段，其中也包括考试这一手段，但不是频繁使用，只在

必要时使用，而且重在利用评价信息。

根据现代教育评价的这一特点，一是要全面地评价学校的办学水平；二是要全面地评价学生质量。在评价学生的时候，既要评价知识、技能、能力，又要评价思想品德、兴趣爱好，还要评价身体发育状况和心理品质等。总之，要对学生的质量进行综合性的评价，不能只评价认知目标。

## 2. 教育评价目标的一致性

这里所讲的教育评价目标的一致性，有两层含义：

第一，是指评价目标要与国家的教育目标和学校的培养目标相一致，因为评价的最终目的就是为了全面地实现国家的教育目标。

第二，是指评价目标要与教学目标相一致，因为依据国家规定的教学计划而编制的教学目标体现了人才培养规格，这是现代教育评价的第二个特点。这一特点充分地体现着教育评价的根本目的。

评价目标与教育目标或教学目标保持一致，这就意味着它们在价值取向上是没有分歧的；如果不一致，这就意味着有分歧，两者存在着脱节的地方。在实施教育评价的时候，如果出现后者，即评价目标与教育目标或教学目标相脱节，那么，这种评价，不论评价指标体系怎么周全，也不论评价方法选择如何好，都是无济于事的，这就是说，这时的评价是难判断教育或教学成果的。

教育评价目标的一致性程度如何，至少有三种因素在起作用：

法律法规约束；

评价理论指导；

价值观的作用。

例如，评价一所学校的好坏，就有两种不同的价值观：一种

是把片面追求升学率作为评价目标，另一种是把全面提高人的素质作为评价目标。显然，前者使评价目标与教育目标相脱节，后者使评价目标与教学目标相统一。因此，在价值取向上，必须把国家规定的教育目的作为教育评价的根本指导思想。

根据现代教育评价的这一特点，在进行评价目标分析时体现两种目标的一致性。就评价学生的质量而言，应从德、智、体等几方面去分析评价目标，而且还应依据不同层次的学校年级分别作进一步的分析，以使评价的具体目标与学校的培养目标和教学目标相统一。从纵向看，这种一致性要体现在层次水平上；从横向看，这种一致性要体现在因素要求上。例如，德育的评价、智育的评价、体育的评价等各个单项评价，都要体现这一特点，以使教育评价能够沿着正确的轨道发展。

### 3. 教育评价机制的指导性

教育评价的目的在于改进。为了实现这一目的，无疑须建立某种运行机制，这就要求在评价的过程中加强指导。因此，指导性就构成现代教育评价的第三个特点。

根据现代教育评价的这一特点，在评价的过程中要给予必要的指导，帮助被评价者改进工作或学习的状态，使其适时地调节自己的行为，以便更好地完成既定的目标。但是，以往的教育评价一般都强调总结性，判断评价对象最终实施目标的结果。片面强调终极结果的弊端是，虽然在评价的过程中发现问题了，但却得不到及时的纠正，以致给工作或学习带来不应有的损失。因此，现代教育评价要求评价者加强评价指导，以便充分地发挥教育评价在教育管理中的作用。

为了有效地进行评价指导，必须做好诊断工作。要达到评价指导的目的，就要发挥评价的诊断机能，对教育的症状进行诊断与分析，及时进行指导。同时，还要发挥评价的反馈机能。也就是说，进行评价指导，必须把有关信息及时反馈给评价对象，让

他们了解问题的症结，指导的针对性就增强了。强调这一点的目的在于发挥被指导者的主观能动性，使内因与外因结合起来，以便取得较好的指导效果。

从评价的结果上看，评价就是要抓住评价对象，按照评价标准判断其达到既定目标的程度。但是，加强评价指导，是重在评价过程的。这里应该注意两种情况：

第一，在评价的过程中，经过诊断和分析，如果确认评价对象是沿着教育目标运行的，并且效果也好，就应予以强化，使其更好地去完成既定目标。

第二，如果确认评价对象有偏离教育目标的现象或其他重要问题，就要适时地加强指导，帮助加以修正。

上述两种情况，均属于教育评价的指导。

从这个意义上看，教育评价的指导过程，也是帮助评价对象实施目标的过程。由此可见，指导目标和评价目标又具有一致性。保持这种一致性，会使教育评价的主体与客体形成一种合力，共同谋求既定目标的实现。

#### 4. 教育评价客体的自律性

要搞好教育评价，就要注意发挥两个积极性，即评价主体和评价客体的积极性。评价客体，特别是人作为评价客体，并不是完全处于被动的接受评价的地位，而是需要发挥主观能动性，自己对自己作出评价。这是现代教育评价对评价客体提出的要求。

在评价的过程中，评价对象作为行为主体，进行自我评价或内部评价，这种现象可以称作评价客体的自律性。一般来说，对个体称自我评价，对团体既可称自我评价又可叫内部评价。重视自我评价，这是现代教育评价论的一个基本观点，也是现代教育评价的一个重要特点。

自我评价与他人评价相比，从效果上看，自我评价优于他人