

教育评估文集

空军指挥学院训练部

教育评估文集

张晓天 选编

空军指挥学院训练部

一九八七年一月

目 录

一、教育评估基本理论与方法

- 1、教育评估的几个理论问题……………李 放 (1)
- 2、教育评价的发展及评价模型的演变……………龚伟民 (5)
- 3、评估的过程、组织与方法……………陈玉昆 (13)
- 4、教育评估对教育系统的反馈调节作用……………孙宝志 (27)
- 5、如何检验一次教育评估的信度与效度……………林鸿剑 (33)
- 6、教育评估量化问题初探……………黄达凯 (38)
- 7、教学评比中的数学方法……………吴伟陵 (49)

二、院校评估

- 8、高校教育评估综述……………张永康 饶乐三 (61)
- 9、评估高等学校工作状态的积分法……………北京航空学院高教研究室 (66)
- 10、关于高等学校的绩效评估……………大连工学院高等教育研究室 (76)
- 11、高等学校教育评估几个理论问题的探讨……………
……………湖南大学教育评估研究组 (80)
- 12、国内研究高等工科教育评估制度的情况综述……………傅雄烈 (88)
- 13、关于制定高等师范专科学校教育水平评估方案的设想……………康永祥 (94)
- 14、略论普通教育的评价……………傅亚宾 张玉田 (99)

三、学科(专业)、课程评估

- 15、论高校重点学科评估二十一条指标……………陈士衡 (106)
- 16、专业评估的尝试和探讨……………曹善华 (107)
- 17、学科(专业)三级评估法……………哈尔滨工业大学 (115)
- 18、高等院校学科评估方案……………湖南大学教育评估研究组 (123)
- 19、高等工程教育学科评估和方案……………陆跃峰 (132)
- 20、机械制造工艺与设备专业(学科)办学水平综合评估指标体系、标准与
方法……………机械工业部教育局 (139)
- 21、专业基础课教学质量评估探讨……………刘延鹏 黄正瑾 (184)
- 22、理论力学课程教学质量评估浅议……………白振林 (191)
- 23、试谈高校《政治经济学》课教学质量评估的几点意见……………肖 飞 (200)

- 24、“电厂热能动力及其自动化”学科的初步自我评估……………南京工学院(204)

四、教师、学生评估

- 25、教师水平多层次结构综合评判……………陶向新(215)
26、教师教学质量语言化综合评定的研究……………殷洪义 李祖培 吴成烈(221)
27、发动学生评价教师教学质量的初步尝试……………郑斯雄 陆曜南(233)
28、对高校数学系学生学习质量评估初探……………唐国华(239)
29、学生德智体定量综合评价方法初探……………张鲁阳 李文库(246)

五、单项评估

- 30、浅议高等学校德育质量评估的原则和方法……………张 宁(250)
31、高等院校科研工作评估指标体系探讨……………刘安产(256)
32、硕士学位质量评估因素……………陶永德 简耀光 周 鸦 燕壮烈(263)
33、《毕业设计(论文)》课及教学质量的评估……………陈子敏(266)

六、国外教育评估

- 34、美国高等教育的评估……………Frank. G. Dickey(275)
35、美国工程教育计划的鉴定准则……………(279)
36、加拿大BC大学的评价制度……………(293)
37、高等院校教学效果的评价标准和指标……………
……………卡济柏日·杰涅克 亚努什·格尼捷茨基(303)
38、日本高等教育评估制度……………吴忠魁(316)
39、大学教学工作评估项目……………丁有豫摘译(323)
40、大学课程的评价……………(328)
41、课程评价检查表……………(335)
42、课程评估指南……………(339)
43、美国教师评价综述……………(344)
44、美国教师评价的理论与实践……………(349)
45、关于教师教学工作的评价……………詹姆斯·奥汉伦 林恩·莫尔斯坦森(335)

教育评估的几个理论问题

李 放

辽宁省高等教育评估研讨会理论研讨组曾围绕着教育评估的一些理论问题进行了讨论。现将会上发表的意见和看法，就我个人的理解和领会汇总如下：

一、教育评估的意义

研讨教育评估的意义，首先遇到现在能不能搞教育评估的问题。对此，有的同志担心，在教育理论尤其是高等教育理论研究的还很不够的情况下进行教育评估，是否能够取得预期的结果，会不会影响教育改革；不如把注意力放在高等教育规律研究上。我们认为这种担心虽然有一定的客观事实和历史因素的根据，但是教育评估却是早已有之的古老事物，现在仍在继续进行着，只不过是没那么自觉，也不是系统的。根据教育史材料提供的信息，我们可以认定，教育评估思想的渊源起于中国，世界上最早有教育评估活动的也是中国。在战国时期问世的《学记》中，就已经谈到考试和考核制度问题，所谓的“比年入学，中年考校，一年视离经辨志，三年视敬业乐群，五年视博习亲师，七年视论学取友，谓之小成。九年知类通达，强立而不反，谓之大成，夫然后足以化民易俗，近者说服而远者怀之。此大学之道也”。这就是明证。所以近人梁启超说：“学校之制，唯吾三代为最备”，诚哉斯言不算为过。隋唐以前，魏文帝定九品官人之法，用以品第人才，成为一种取士制度，隋唐以降又实行科举制度，以九经取士，吏部试通过便可委以官职。因而，“学而优则仕”便成了学校教育的目的。晚清时，洋务派和维新派虽然他们的政见不尽相同，但最终还是导致废科举，兴学堂，并在政府部门设置视学官，专事巡视京外学务，备有《视学官章程》。中华民国教育部也曾公布视学章程，设视学，秉承长官之命，掌学务之视察。在我国建国以来，教育评估的实践活动也是很多的，具有比较完备的考试制度，工作计划、检查、总结制度；教师职称评定、晋升制度，等等。尽管这些还不能称之为教育评估理论和科学的评估活动，但是在不同程度上确是反映了教育评估的思想，并具有一定的教育评估理论的因素。这是值得研究和总结的可贵经验。

现在，我们正在研究和探讨教育评估问题，使之逐步走向科学化、理论化、系统化、制度化，以充分发挥其比较、激励和“奖优促劣”的作用。在这个意义上说，实行教育评估一定会有利于贯彻“教育面向现代化，面向世界，面向未来”的战略方针；有利于教育体制改革，为社会主义建设多出人才，出好人才；有利于提高学校的办学水平和教育质量；有利于校际之间的交流和促进。所以，我们可以有根据地说，现在实行教育评估是势

在必行的,《教育体制改革的决定》中提出“教育管理部门还要组织教育界、知识界和用人单位定期对高等学校的办学水平进行评估”是完全适宜的,正确的。至于对教育规律的研究,如同其他科学一样,已经认识和掌握了一些,但是科学是无止境的,人们还在继续揭示着教育规律。而教育评估则是一个综合性的问题,不仅要借助于教育理论,而且要借助于心理学、统计学、政治经济学、管理学,以及其他许多学科的研究成果,才能综合地对教育现象进行评估,也才能丰富和发展教育评估理论,就是说,对教育规律的认识和揭示虽然不够充分,但已具备了开展教育评估的基础和条件;自不待言,正确的教育评估对教育改革具有重大的推动和指导作用,而教育改革在教育评估的影响下,才能更好地坚持正确的方向,取得积极的社会效益。因此,不应该把教育评估和教育改革对立起来。

二、教育评估的科学含义和概念定义

教育评估作为一门学科,其理论体系是在美国批判测量运动中形成的,并逐渐取代了教育测量。它的理论和基本思想,就是注意教育效果的价值观,分析教育应达到的目的,根据教育目标来评价教育效果,并认为有了教育评估就能使教育措施为实现理想教育目标起保证作用。但是,由于人们看问题的角度、方法不同,对教育评估的含义和定义,便出现了几种不同的理解:有的同志从教育管理学的角度看教育评估,就把它看作是一种科学管理的手段;也有的同志从当前科学发展的角度来看,认为教育评估将会形成和发展成为一门独立学科,应该有它自己研究的对象、任务和方法;还有的同志从系统论的角度强调,首先必须把教育评估的要素弄清楚,才能比较容易或比较科学地阐述教育评估的涵意。唯其如此,对教育评估的科学含义和定义,也就作出各自不同的回答:

一种看法认为,教育评估是根据评估理论,确定评估目标,并以此对各项教育工作从数量、质量、效益等方面进行评估的一种科学管理手段。

另一种看法认为,教育评估就是以马克思主义为基础,运用系统科学的方法,对学校办学水平进行综合评价和估量,以促使评估对象趋向最优化。

还有一种意见认为,教育评估应该是以事实为依据,以理论为指导,按着科学确定的指标,对教育整体的数量和质量进行综合的评议和估价。

第四种意见认为,教育评估应该是以马克思主义教育理论为指导,根据各级各类学校及其各项教育活动的目标函数,运用科学的方法来评定和估量教育现象的社会价值,以促进教育对象趋于最优化的一种行为。

这四种意见各自都有其合理因素,在不同程度或方面反映了教育评估的实质,它们随着研讨过程的不断深入都有所变化,并将在实践中进一步完善。一般地说来,会上多数人比较倾向或赞同最后一种意见。随着对教育评估问题研讨的深入,总会形成一个更科学更全面的概念定义。

三、教育评估的理论基础

教育评估作为一门学科或理论体系,它不仅有自己的独特的研究方法,还要有一种普

遍的科学认识的方法论，构成它研究问题的原则基础。这是毫无疑义的。问题在于用什么作为它的理论基础。在这个问题上有两种不同的看法：一种认为系统论、信息论、控制论，即所谓的“三论”应该作为教育评估的理论基础；另一种看法认为，必须坚持以马克思主义唯物辩证法作为它的理论基础。这两种看法在研讨过程中逐渐趋向于一致。关键是如何看待唯物辩证法与“三论”的关系问题，以及它们与某一门学科的研究方法的关系问题。对这三者究竟应该怎样看待和处理，是当前教育理论界一个很重要的新课题。这个问题的解决具有很大的理论意义和实践意义。

我们认为马克思主义哲学是各门科学的方法论基础，每门科学都有它自己的研究方法，当它运用自己的研究方法研究这门科学对象时，就必须依靠马克思主义的方法论的指导，而由此所取得的认识成果，又作为思想材料来充实马克思主义。这是哲学和各门科学之间的必然联系。由于现代科学的迅猛发展，开始在自然科学及其研究方法领域里出现了“三论”，并作为横向科学贯通在一些学科之中，继而又把“三论”逐渐引进社会科学领域，在研究方法方面起着积极的作用。但是，它没有取代马克思主义的方法论，尽管有人说：“由于‘三论’的出现，使传统认识论中关于实践对认识的单一决定模式进到现代认识论的实践和认识既先后相互作用、又能相互改造的模式”，但是人类的认识说到底是源于实践的，完全脱离实践的认识是没有的；认识最终要受实践的检验。至于说实践和认识的辩证统一关系，这是马克思主义认识论的实质。可见，“三论”不仅没有超过马克思主义，也没有发展马克思主义，只能说它对马克思主义方法论有所丰富和补充，它是马克思主义方法论和各门科学的研究方法的一个中间层次，起着枢纽和桥梁作用。因此，在这个意义上说，我们研究一门科学或研究一个事物，当然可以采用“三论”的方法。但是，不能把它作为指导思想，也不能把它作为理论基础。它们的理论基础和指导思想只能是作为科学的世界观和方法论的马克思主义哲学。就是“三论”本身也不能脱离或否认这一点。正如毛泽东同志所说的“指导我们思想的理论基础是马克思主义”，这是一条颠扑不破的真理。

四、目标评估和过程评估

教育评估对象除了目标评估和过程评估，还有条件评估和贡献评估。但是，目标评估和过程评估是基本的、主要的；有时它们也可以把条件评估和贡献评估包容进去。当然，条件评估和贡献评估也可以单独进行。譬如，目前有一些地方要办专科学校或变专科学校为本科院校，可是它究竟应具备哪些条件方可批准？这就需要由条件评估来加以衡量和解决。但主要的还是目标评估和过程评估，因为它们是经常的、大量的、是具有决定意义的。

所谓目标评估，我们把它理解为主要是反映效果、效益、教育质量和办学水平的一种评估；所谓过程评估，主要是指办学条件、学校的工作过程和主观努力的程度而言的。也有人把办学条件列入目标评估，这似乎有些不妥。因为条件本身既不是效果和效益，也不是质量和办学水平，仅仅是构成它们的一个客观因素。

在过程评估里，既有客观物质条件，也有人们主观能动性方面的条件，两者结合起

来，形成过程评估。目标评估和过程评估的关系，从理论上讲，教育是一种有目的、有计划、有组织的一种活动，它是一种运动过程，必然产生一定的效益和结果；没有无运动的结果，也没有无结果的运动，二者是辩证统一的。从实践上说，当前的教育评估应着力于目标评估，即评估实现教育目标的那种教育效果。并逐步把目标评估和过程评估有机地结合起来。教育评估正是在这点上区别于历史上曾经出现的教育测量。二十世纪初期，教育测量在美国盛极一时，堪称异军突起、所向披靡。它的理论根据是桑代克的著名信条，即所谓的“凡是存在的都有数量，凡有数量的东西都可测量”。由此逐渐形成测量运动。但是，在实践中暴露出这种方法的局限性，它不能反映出人们的社会态度，实际技术、创造兴趣、鉴赏能力等等，而被三十年代兴起的教育评估所取代。教育评估既把主要注意力放在对于效果、效益和质量方面的评估，又不排除定量方面的研究和评定，这就是它的生命力之所在。因此，目标评估应该是评估的主体，起决定作用。而过程评估则是从属于目标评估，是为目标评估服务的。这里应该指出的是，目标评估和过程评估在实践上有着密切联系；我们进行目标评估，同时也不可能脱离过程评估完全忽视实现目标的条件，同样，过程评估也不可能完全不考虑效果和效益问题，二者是对立统一，有机结合的。但是，在一定的时期，一定的问题上，它们可以有所侧重，或是侧重目标评估，或侧重过程评估；侧重并不是排它性的，而是在构成矛盾的统一体中居于主导地位。比如，两所学校培养人才的质量相同或相近，如果只把注意力放在目标评估方面，而不联系和考虑过程评估，那么所得出的结论就难以客观地反映它的真实性。因为有的学校条件好些，有的学校条件差些，在条件有所差异的情况下，这两所学校都达到一个相同或相近的目标水平，显然，条件差的学校工作是好的、是优化的；反之，条件好的学校，工作上就有差距，人们主观能动性还发挥得不够充分。所以，我们认为，应该把目标评估和过程评估很好地结合起来，并有所侧重，从而正确地发挥教育评估的作用。

以上谈到教育评估的几个理论问题，虽然不是教育评估理论的全貌，但它们却是教育评估理论的重要内容和组成部分，我们在这里所谈的意见，只是“抛砖引玉”，以期把教育评估理论问题的讨论引向深入，并能指导和服务于教育评估的实践。

教育评价的发展及评价模型的演变

龚伟民

评价学是一门古老的艺术。如果中国古代封建社会特有的科举制度可以算作最原始的评价形式，它至少已有几千年的历史了。评价学是一门新兴学科。建立在心理测量基础上的现代评价学，产生于19世纪初，迄今不到二百年。从60年代到现在短短的二十多年时间，随着评价理论不断丰富，评价方法的不断完善，评价已成为一门完整的学科，越来越受到人们的重视。以美国为例，据1984年统计，美国大学目前共有117个侧重于评价的专业点，可以授予硕士、博士以至博士后学位。其中约1/2属于教育专业，1/3属于心理专业，还有一些分别属于社会学，社会工作等其它专业。又据1980年统计，美国现有专门评价杂志7份，部分涉及评价的各类杂志几十份。每年出版的有关评价的专著更是不计其数。美国有两个全国性的评价组织。一个称之为“评价网”（Evaluation Network），成立于1975年；一个称之为“评价研究协会”（Evaluation Research Society）成立于1976年。这两个组织的会员具有广泛的代表性，除了来自教育部门以外，还来自政府机关、工程技术、商业、医疗卫生等各个部门。它们向不同专业的评价人员提供讲坛，使她们有机会在保持本专业特点的情况下，对评价理论和评价方法发表自己的看法，相互学习，共同研究。除了全国性组织，许多州、地区也有自己的评价团体。这些团体对政府决策起着重要的影响。可见，作为一种社会现象，一种人类活动，评价已是现代社会所不可缺少的了。教育、卫生、社会福利等公共事业的每项活动，无一不需要以评价支持其发展、改进其不足，以评价对其作出价值判断。为了对现代评价的理论和方法有个大概的了解，我们将评价的发展历史及评价模型的演变作个简单的回顾和介绍。

研究评价学的发展历史，可以从不同的角度着手，比较评价模型仅是其中的一种而已。所谓“评价模型”，与通常所说的数学模型、经济计量模型等精确模型相差甚远，有人甚至反对使用这个字眼。但是，作为一种抽象，评价模型集中地体现了评价者所依据的评价理论，所希望达到的评价目的，从而也大致确定了评价所使用的方法。所以，它是评价学的重要研究课题之一。

19世纪以来的评价学发展大致可分为以下四个时期：

- （一）19世纪初——20世纪30年代，或称之为萌芽期。
- （二）20世纪30年代——50年代末，或称之为泰勒期。
- （三）50年代末——70年代中，或称之为发展期。
- （四）70年代中——现在，或称之为成熟期。

除第一时期外（尚未形成完整的评价体系），我们将以当时最具代表性的评价模型为基本线索，对这几个时期的评价理论及其方法进行比较，以窥其全貌。

（一）萌芽期：19世纪初——20世纪30年代

在教育评价发展初期，教育测验和教育测量是评价的主要手段。1845年，美国波士顿的教育工作者，首次使用对学生进行测验的方法，来评价学校效果。测验形式一改以往单纯的口头回答形式，转向使用较为客观的书面论文形式，从此开始了长达百年的使用标准化测验的历史。1887年到1898年，被誉为“教育科学研究之父”的约瑟夫·赖斯（Joseph Rice）为了解学校时间是否被有效地利用，设计了一套学业成绩测验。学生在测验中取得的成绩，成为他判断的标准。莱斯发现，有些学校每星期花在拼写单词上的时间多达200分钟，而有些学校则少至10分钟，但这些学校的学生在测验中取得的成绩并无显著差异。这个结果说明，某些学校的拼写教学需要大大改进。

1904年，阿尔弗雷德·比奈（Alfred Binet）在法国公共训练部的授意下，设计了一套心理测验，其主要目的在于从普通班级中发现那些智力发展迟钝的学生，这套测验为以后智力测验的产生打下了基础。第二次世界大战大大刺激了智力测验的运用和发展。由于战争，一方面要求部队尽快地从大量的应征人员中挑选出合格的、有能力的学生，一方面又要求学校尽快地采取各种手段培养那些有天赋的学生，所以，各种智力测验应运而生。

纵观这一时期的评价活动，主要有以下几个特点：

1. 评价和测量是两个可以互换和通用的概念，实际上，测量一词较评价更为常用。评价活动一般总离不开测量。

2. 评价和测量都与科学探究方法紧密地联系在一起。19世纪到20世纪这段时间，科学探究方法取得了很大的进展。1843年，约翰·斯图亚特·米勒（John Stuart Mill）提出，研究社会现象，应该采取科学探究法而不应是哲学的，或神学的方法。这种倾向在很大程度上受到达尔文名著《物种起源》的影响。莱斯在研究学校效果时，也是以充实的测量数据为依据，后来泰勒领导进行的“八年研究”，所采用的方法，基本也是诸如建立控制组，对被试体施行事前和事后测量等科学实验法。

3. 评价和测量的主要焦点集中于被测对象的个体差异。这也与当时自然科学的研究水平有关。从教育观点来看，它显然忽视了许多教育工作者应当关心的重要问题，因而是很不全面的、狭窄的。

4. 评价和测量与学校的教育方案，课程很少有直接联系。测量能够揭示某些学生个体之间的差异，但单纯的测量（包括学生学业成绩测验和心理测验），远远不足以说明造成这种差异的教学方案和课程，从而无法解决教育存在的问题。

5. 为了使测验结果相互可比，当时均采用常模参照形式的标准化客观测验。建立可靠的常模需要经过大量的抽样和严格的可控试验，这是一件艰巨的工作，在当时并未很好地解决。

6. 当时的教育界受工业界流行的“科学管理”思潮影响颇大，尤其是第一次世界大战以后，教育界倾向于采用企业家管理工厂的方法来管理学校。在这种情况下，科学化、

标准化的测量成为评价的主要手段是毫不奇怪的。

以测量为基本手段的评价徘徊了100来年，终于在20世纪30年代有了重大突破。这一成就无疑应归功于泰勒和他的同事们。

(二) 泰勒期：20世纪30年代——50年代末

拉尔夫·泰勒 (Ralph W. Tyler)，俄亥俄州立大学教授，1932年为著名的“八年研究”负责人。所谓泰勒原则是由他的二项研究组合而成。即1934年发表的“成绩测验设计” (Construct achievement test) 和1945年发表的“课程与教学的基本原则” (Basic principle of curriculum and instruction)。

1. 成绩测验编制

1929年，泰勒应俄亥俄州立大学教育研究所所长的邀请，参与帮助改进学校的测验方法。学校认为，改进测验和测量方法，也许能提高教学质量和提高学生的积极性。

当时的的问题是，以往测验仅考察学生对课本规定内容的机械记忆，并不考察学生是否掌握了运用课本知识的复杂行为和技巧。所以，虽然学生能在测验中获得较好成绩，但却缺乏实际动手能力和解决问题的能力。泰勒认为，一份好的测验，不是简单地从教科书中挑选一部分内容，改写成问题形式，要求学生对此加以重复，而是应该以一定的教育目标为指导。当然，这样编制测验要比原来的方法困难得多。所以，泰勒要求教师首先找出课程目标，然后根据他们所希望的，学生应该掌握的内容、方法，用行为将课程目标表述出来，制定出一系列教育目标。他认为，只有这样编制的测验，才能够不但检查学生的记忆能力，还能检查学生对现象的理介以及实际运用知识进行推理和应用的能力。

2. 课程理论

30年代，美国中学课程改革的主要问题是：学校的教师感到，由于严格的大学入学要求和州对每个学校规定的认可标准，学校很少有修改课程的可能性。学生往往在小学阶段有较多的自我发展的机会，而进了中学则被抑制了。另一方面，由于当时的经济萧条，许多青年因为无法就业而只能进学校，但学校的课程又无法吸引这批青年人。所以，满足经济萧条时期青年的要求并为他们提供较多的自由发展机会成为改革的焦点。在“进步教育协会”的组织下，在全国范围内挑选了30所中学，让他们在不考虑大学入学要求和认可标准的前提下，花八年时间进行理论改革。在这个基础上，泰勒和他的同事设计了“课程和教学的基本原则”。这些原则是：

- (1) 确定教育目标，包括那些是学生应该学习的思维、情感和行为方式。
- (2) 确定达到教育目标所必备的学习条件。
- (3) 最合理地组织学习条件。
- (4) 不断对方案的效果进行评价。

总之，泰勒原则的评价过程可归纳如下：

- (1) 通过研究实际需要和征求专家意见，选出备选目标。
- (2) 经过若干筛选过程后（如哲学、心理学、实验等），确定可用目标。
- (3) 将选定的目标用有关的术语加以表述。
- (4) 找出这些行为表现的条件和情景。

- (5) 确定满足客观性、可靠性、有效性三个标准的测量方法。
- (6) 应用这些方法（通常是事前、事后测验）检查行为变化。
- (7) 根据检查结果对假定作出判断，并说明原因。
- (8) 修改方案，重复循环过程。

泰勒的评价原则是评价的一大进步。在这之前，评价主要用于判断学生个人的特征，而他首次将评价运用于课程改革之中，评价成为课程改革的一个不可分割的组成部分。评价的依据和中心是目标，而评价的结果又不断地修正、改进着目标。从目标的确立到评价结果不是直线型的单向过程，而是周期性的循环过程，泰勒第一次明确地将测量与评价作了区分，前者只是后者的一个工具，至此赋予评价以新的意义。

泰勒模型结构紧凑，逻辑严密，简明易懂，但同时也暴露出以现代眼光看来所不足之处。首先，它没有提供评价目标本身的方法。也许某些目标是通过筛选的，但模型并未说明这个筛选网是如何组成的。筛选网应该包含哪些因素。当评价者怀疑评价目标，或者在评价过程中发现评价目标有所不足，他应该怎么办？

其次，模型没有提供能够用来判断评价目标与评价结果之间差异（或矛盾）的标准。这个差异要多大才需要引起重视？这个差异有什么意义？模型假定这个差异的存在，并以此收集有关的信息。显然，当假定与事实不符时，所收集的信息就可能是没有多大实际意义了。

第三，模型将注意力集中于事前、事后的测验设计。泰勒认为，一个行为应该在受到“处理”之前和之后经过两次测量。这种完全以自然科学的实验作为社会科学研究的方

法，始终为人们所争议。

泰勒模型从30年代起，一直占据着评价领域的主导地位，这种局面到50年代末才被打破。

三、发展期：50年代末——70年代中

1957年10月4日，苏联发射了第一颗人造卫星。这一事件大大震动了美国朝野，从而改变了美国教育的面貌。美国教育界不甘心落后，开始进行大规模的课程改革，成百万的金钱用来改造和编制新课程，如国家科学基金会资助的PSSC物理课程、SMSC数学课程和由美国教育总署资助的英语、社会研究等项目。既然投入了如此大量的资源，对因此而产生的结果进行评价就成为至关重要的了。但时隔不久，人们就抱怨评价在促进课程改革方面没有起到应有的作用。1963年，克龙巴赫（Cronbach）在一篇题为《通过评价改进课程》的文章中，对当时的评价提出三点质疑，他认为：

(1) 评价者不仅应关心课程制定者规定的目标，检验这些目标达到的程度，更应关心谁是决策人？他们作了些什么决策？他们按照什么准则作决策？等等。显然，评价的中心不应仅是目标，而更应是决策。

(2) 在课程发展过程中进行评价比在课程结束后进行评价更有助于课程的改革。所以，评价应将重点放在教育过程之中，而非教育过程结束之后。

(3) 对不同课程各自的特点进行评价比将它们简单地相互比较更有意义。从而对评价的理论、技术提出了新要求。

1967年，斯克里芬（Scriven）发表了评价史上具有深远影响的《评价方法论》。在这篇文章里，他指出，评价不但在理论上，而且在实际运用中都很不完善。他对以下几个评价概念作了区分：（1）形成性评价和终结性评价；（2）专业性评价和业余性评价；（3）仅测定目标达到程度的评价和不仅测定目标达到程度，而且关心这些目标是否值得达到的评价；（4）比较性评价和非比较性评价，等等。

1968年，艾斯纳（Eisner）对目标的概念进行了评击。他认为，对教育本质的不同理介，可以造成对目标的不同陈述。他提出三个假设：（1）工业化假设；（2）行为化假设；（3）生物化假设。泰勒的教育目标是前两个假设的产物，而教师希望达到的目标则根据第三个假设而来。所以，行为和目标之间的差异可以评价泰勒的目标，但都不适用于教育实际。这就是说，评价方法必须有个根本的改变。

可见，实行了20余年的泰勒模型在苏联卫星的冲击下动摇了。总的来说，这段时间评价理论的主要变化是：

- （1）目标的价值结构受到挑战，目标本身成为判断的对象。
- （2）评价者不应仅是资料的收集者，更应是目标的判断者。
- （3）以目标为中心的评价方式不再是唯一的了，其它评价形式相继出现。
- （4）产生了一大批新的评价理论，人们对评价的功能有了更广泛、更深刻的认识。

在这一时期内出现的评价模型大约有40多个，其中最著名的有：表象模型（Counter-raue Model），CIPP模型，目标自由模型（Good Free Model），行家鉴定模型（Connuisseurship Model）等等。

CIPP模型是斯塔弗尔比姆（Stuggebean）于1966年提出的。这个模型曾被美国许多院校和机关团体所采用，在学术界也曾引起过广泛的讨论。与传统模型不同，CIPP是一个以决策为中心的评价模型。它不要求有关目标的信息，而只要求了解作出了什么决策？谁作出的决策？怎么作出的和根据什么作出的，等等。CIPP是四种不同类型评价的缩写，分别表示背景评价（Context），投入评价（Input），过程评价（Process）和结果评价（Product）。斯塔弗尔比姆和他的同事们发现，在按泰勒模型对某个方案进行评价时，有许多难以解决的困难。因为泰勒模型要求评价者找出每个评价目标的行为目标，采用适当的方法测量学生的表现，从而比较测量结果与目标的差异。但实际上，许多方案目标是由方案制定者所规定的，他们对学生可能的表现并不了解。如果方案制定者和实际评价者对说明评价目标的行为理介不一致时，评价就很难进行下去。此外，泰勒模型的评价结果通常出现在方案之后，不能有效地提供反馈信息。

考虑到泰勒模型的种种不足，斯塔弗尔比姆认为，应赋予评价以新的定义，即“评价不应局限于确定目标的达到程度，而应成为为决策提供信息的过程”。为了达到这个目的，他对方案执行过程中的决策类型作了一个简单的划分，每个决策类型分别对应一种评价类型，从而组成CIPP评价体系。这些组成是：（1）预期结果的决策。为这类决策提供信息的是背景评价；（2）预期方法的决策。为这类决策提供信息的是投入评价；（3）实际方法的决策。为这类决策提供信息的是过程评价。（4）实际结果的评价。为这类决策提供信息的是结果评价。在整个模型中，背景评价是个不断循环的往复过程，而投入、过程、结果评价则在背景评价提出需要，发现问题，并要求予以解决的时候进行。

CIPP模型是第一个摆脱传统目标局限的评价模型，它更好地反映了社会对评价提出的新要求，也全面、系统地反映了评价对象的全貌。但是，作为一个以决策为中心的评价模型，它在许多方面并不是行得通的。如：对决策者理性思维的假设；对决策过程公开化的要求；对所有决策层次的可识别性等等。

表象模型 (Counterance Model)

以目标为中心的泰勒模型尽管受到很多责难，但仍有人坚持这个原则，并在此基础上以发展和充实。斯塔克 (stake) “表象模型”就是其中的代表。

斯塔克提出的模型包括两个数据矩阵：描绘矩阵和判断矩阵。每个矩阵分为二列：意图和观察组成描绘矩阵；标准和判断组成判断矩阵。每个矩阵又分为三行：前提、执行和成果。见下图：

意图	观察	前提	标准	判断

评价者首先确定意图列中的三行。这里的前提指方案实行前的条件，也就是通常所说的投入。执行指方案实行中可能遇到的情况。成果指期望方案取得的结果。与泰勒模型不同，斯塔克不仅强调成果的目标，同时也强调与方案有关的背景条件的目标。

接着，评价者收集观察到三行的信息。对应于意图到的三种目标，观察到收集的信息分别显示实际情况与预想目标之间的差异。泰勒模型只强调成果与目标之间的差异，由于在斯塔克模型中还包括了前提和执行的情况，那么，当预想的成果没有获得时，就能够对方案的前提和执行作进一步的假设和检验，从而找出失败的原因。

在标准一列中，斯塔克使用了两类标准：绝对标准和相对标准。前者由某一特定的集团建立，后者由观点不同的集团建立。这有点象测量中的标准参照系统与常模参照系统。

最后的判断列，主要用于解释观察与标准之间的差异。判断的一个重要功能是为不同的标准加上不同权数，以至最终能对方案的价值作出结论性的说明。

在表象模型中，斯塔克将目标的概念扩大了，它不但指成果，同时也包括前提与执行。他第一次将判断视为评价的主要任务，将描绘和判断作为评价中不可分割的两个组成部分。他建议用两类标准作为判断的基础，以及从前提、执行、成果三个层次分析方案成败的原因。

表象模型的缺点在于没有说明如何获得标准，无论是绝对标准还是相对标准。它的基本假设未脱离泰勒模型，即社会价值观念的收敛性（区别于应答模型的发散性），和检验手段的科学实验性。

(四) 成熟期：70年代中——现在

各种评价模型的出现，反映了社会对评价的不同要求和期望。作为一个整体，这些模型比起泰勒时期单一以目标为中心的评价来，无疑是一个很大的进步。明确价值和标准已越来越迫切，判断已成为评价不可缺少的部分。不仅如此，人们也开始对评价所采用的科学方法提出了质疑。评价不但从理论上，而且从实验上都开始发生重大变化。

这里，出现了几个引起人们思考的问题：（1）评价越来越成为一种政治现象，但很少有人了解评价究竟如何受到不同政治倾向集团的影响。（2）评价越来越被人的因素所左右，但很少有人了解人的因素对成功地收集信息究竟有多大影响。（3）要求意见一致，观点统一，即价值观念收敛性的模型越来越难以满足一个多层次的复杂社会的需要。正是在这种情况下，新的评价模型出现了。其中最具代表性的是所谓“应答模型”（Contentance Model）。

应答模型是由斯塔克首先提出，由古巴、林肯（Guba & Lincoln）等人进一步发展而形成。

斯塔克认为，评价一个方案可以有不同的方法，没有一种方法是绝对正确的。但要使评价产生效用，有一点则必不可少，即评价应该向听取评价结果的人提供他们所关心的信息，评价者应该充分了解那些人所关心的问题。他还认为，如果一个评价，它牺牲某些测量上的准确性以换取对听取评价人的有用性，它更关心方案的活动而不是方案的内容，它更注意反映与方案有关的各方面人的意志而不仅是一部分人的意志，那么，这种评价就是应答式评价。

古巴和林肯进一步指出，所谓应答模型，就是以所有与方案有利害关系或切身利益的人所关心的问题为中心的一种评价。这些问题可以包括：是否应该削减方案预算？是否应实行责任制？教学目标是否达到？新教学计划比旧教学计划是否更为优越？等等。总之，任何对方案的担心、怀疑、赞成、或反对，均为应答模型所应该为之提供信息的。

应答模型的评价步骤有以下几个：

（1）评价者与一切跟评价对象（如教育方案）有关的人接触，获取他们对评价对象的想法。

（2）根据获取的信息，确定评价范围。

（3）评价者亲自对方案的实施作实地观察。

（4）通过观察，评价者对方案所希望达到的目标与实际上取得的成果行行比较。

（5）评价者对评价应回答的问题进行理论上的修正。

（6）在此基础上评价者设计评价方案。

（7）根据不同的要求，选择不同的收集信息的方法。

（8）对收集来的资料进行加工处理。

（9）将处理过的信息按需要回答的问题分类。

（10）把分类评价结果写成正式报告并分发给各有关人员。

（11）在分类评价报告的基础上对方案作出全面判断。

应答模型采用了与以往评价模型完全不同的评价方法，即自然主义方法。

应答模型的出现，一方面是社会对评价进一步的需要，一方面也是评价这门学科自身发展的必然趋势。

支持并赞成应答模型的评价专家们认为，应答模型是迄今为止所有评价模型中最全面、最有效的。其理由主要有两个：

首先，这种模型不再单纯从理论上出发，而是从关心评价结果的各种观众的需要出发。任何评价，如果其结果不能为观众所接受，那么，它的理论再高明，方法再先进，也

是产生不了任何效果的。

其次，应答模型回答了所有其它模型希望回答的问题，包括目标达到的程度、决策、判断等。它更适合于一个多元的，复杂的客观世界的现实和不同地位，不同观点的评价观众的需要。它的结果是具有弹性和应变的。

我们对评价学发展历史作个简要的概括。在现代评价学发展的近二百年的时间内，大致经历了四个时期：萌芽期、泰勒期、发展期和成熟期。在萌芽期内，测量是评价的主要手段，评价以学生的个体差异为主要对象，很少与教育方案发生联系。在泰勒期内，测验与评价开始分家，测验成为评价的一种工具。泰勒模型是以目标为中心的一种评价形式。在发展期内，泰勒模型受到了抨击，各种以决策为中心（如CIPP），以判断为中心（如行家鉴定）的评价模型相继出现，同时泰勒模型也得到了进一步的发展（如表象模型）。在成熟期内，评价的理论和方法有了根本性的变化。应鉴模型是最著名的代表。

评估的过程、组织与方法

陈玉琨

教育评估是一技术性很强的工作，能否科学地组织评估对评估的质量与结果的可靠性和有效性有着重要的影响。作为按系统程序进行的工作，评估过程大体可分为准备、实施和结果分析三个阶段。就其组织来说，教育评估一般又可分为领导评估、同行评估和自我评估等三种类型。评估的方法从不同的角度可作不同的分类。下面，我们对评估的过程、组织和方法作初步的分析。

（一）评估的过程

科学的评估过程是一个按系统的程序进行工作的过程。一般地，它可分成准备、实施与结果分析与处理三个阶段。

1. 准备阶段

准备阶段是评估具体实施前的预备阶段。凡事予则立、不予则废。这对评估过程来说是必不可少的。准备阶段的质量将直接影响评估的质量。准备阶段又包括组织准备、方案准备和文件准备三个方面。但是，不论是组织准备，还是方案、文件准备，它们都是以特定的评估目标为出发点的。因此，评估的准备工作都必须首先始于明确评估目标，然后再作各方面的准备。

（1）组织准备。组织准备包括成立专门的评估委员会、聘请有关专家，组成专家评审组和设置一定形式的评估办事机构。当然在不同类型的评估活动中，相应的组织形式也是不同的。这个问题，我们将在下面作进一步的论述。

（2）方案准备。方案准备在整个准备阶段中是具有实质性和关键性意义的工作。方案准备主要包括下列任务：第一，明确评估目的和评估所依据的目标。一个特定的评估活动所想达到的目的，对评估方案的准备、评估的实施、乃至对结果的分析处理都有影响，因此，评估的准备必须从明确评估的目的开始，比如，以分清高校教育工作的优良程度为目的的评价，与以衡量高校是否达到了及格标准为目的的评价，显然，在评估的方法与标准上是极不相同的，前者用的是相对评估法，即通过被评对象相互比较而得出评估结论的，而后者一般则用绝对评估法，它是用规定的标准去衡量和判断学校是否达到了应有的水平。因此，为了顺利地开展评估工作，我们首先必须明确评估的目的。评估所依据的目标与评估的目的并不是一回事，它指的是，各类高等院校的办学目标或管理目标。这一目标是作出评估结论的依据。比如，高层次的大学（即原来意义上的重点大学）它的目标是努力办成既能出高层次高质量的人才，又能作高水平的科学研究的“两个中心”，这一目