

国家自然科学基金资助项目

非线性评估的 理论探索与应用

侯定丕 王战军 主编

中国科学技术大学出版社

国家自然科学基金资助项目

非线性评估的理论探索与应用

主 编：侯定丕 王战军

编著人员：侯定丕 王战军 张淑林

吕晓澎 林梦泉 侯富民

赵 瑜 周学军 陈 伟

安 勇 何 俊 杨周旺

张升宇 王惠文

中国科学技术大学出版社

2001·合肥

图书在版编目 (CIP) 数据

非线性评估的理论探索与应用/侯定丕, 王战军主编. —合肥: 中国科学技术大学出版社, 2001. 2

ISBN 7-312-01191-8

I. 非… II. ①侯… ②王… III. 非线性理论—评估—方法
IV. N945.16

中国版本图书馆 CIP 数据核字 (2000) 第 22604 号

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路 96 号 230026)

中国科学技术大学印刷厂印刷

全国新华书店经销

开本: 850×1168/32 印张: 7.75 字数: 202 千

2001 年 2 月第 1 版 2001 年 2 月第 1 次印刷

印数: 1—2000 册

定价: 20.00 元

前 言

评估，亦称评价，是科学管理工作中的一个重要环节。常见的评估工作总是采用加权平均求综合分的方法，这是以评估是线性的为前提来进行的。

我们从多个领域的评估工作实践经验出发，认识到评估的本质是非线性的，不应该一律用线性方法来进行。我们探讨了非线性评估方法及其若干理论问题，如评估指标体系及其简化、对评估者进行的评估分析、评估模型的检验、评估模型的拟合与概念生成型不确定性处理等，有些其实就是元评估所涉及的问题。在本书中，我们结合近年来的评估实践与探索心得，对这些问题进行了论述。

本书可供广大评估工作者（特别是教育领域）、管理科学工作者以及对评估理论感兴趣的大学生、研究生和其他人员参考。

本书对这些问题的探讨，无疑还处在起始阶段，因此不当之处自然难免。但是，我们希望这本书能起到抛砖引玉的作用，大家一起努力，共同促进评估工作科学性的提高。

编 者

2000年12月26日

目 次

前言	I
第1章 评估工作常规方法论	1
1.1 评估的基本概念	1
1.1.1 评估的基本概念	1
1.1.2 教育评估	4
1.2 评估的基本功能	10
1.3 评估的常规方法	12
1.3.1 相对评估法	12
1.3.2 绝对评估法	13
1.3.3 个体内差异评估法	14
1.3.4 数量化方法	15
1.3.5 非数量化方法	17
1.4 评估的基本程序	18
1.5 案例报告 1: 评估指标体系的建立	22
1.5.1 建立评估指标体系应注意的问题	22
1.5.2 Delphi 法	23
1.5.3 重要性指标筛选	29
1.5.4 小结	33
1.6 案例报告 2: 层次分析法的应用	34
1.6.1 层次结构模型	35
1.6.2 成对比较法	37
1.6.3 正逆称方阵	39
1.6.4 应用	50
1.7 案例报告 3: 增列博士、硕士学位授权点评估	53
1.7.1 概述	54

1.7.2	评估目的与评估对象.....	55
1.7.3	评估指标体系、指标内涵及其权重集.....	56
1.7.4	评估过程.....	59
1.7.5	评估结果与分析.....	62
1.7.6	程序设计.....	65
1.7.7	有待探讨的若干问题.....	66
1.8	案例报告 4: 全国 5 个一级学科选优评估.....	68
1.8.1	概述.....	68
1.8.2	评估指标体系、指标内涵及其权重集.....	72
1.8.3	评估方法与技术.....	76
1.8.4	评估结果与分析.....	81
1.8.5	评估的收获与存在的问题.....	86
第 2 章	评估的非线性特征.....	89
2.1	线性评估的定义.....	89
2.2	导致评估非线性的若干缘由.....	91
2.2.1	特殊评估指标的作用.....	91
2.2.2	规模效应指标的作用.....	93
2.2.3	评估对象的和.....	94
2.2.4	评估对象部分指标的数量扩缩.....	95
2.2.5	政策因素所起的作用.....	96
2.2.6	社会反馈.....	97
2.2.7	数据处理.....	98
2.2.8	定性评估.....	99
2.3	非线性与评估工程.....	100
第 3 章	非线性评估方法及案例.....	102
3.1	门坎—线性式法.....	102
3.2	阶梯型模型法.....	104
3.2.1	阶梯型模型法.....	104
3.2.2	案例.....	106
3.3	字典顺序式法.....	113
3.4	层位模型法.....	114

3.5 非线性综合评分式法	117
3.5.1 生产函数式	117
3.5.2 对数求和式	119
3.5.3 加乘总和式	120
3.6 理想解法	123
3.7 一种成对比较排序法	128
3.8 模糊综合评判	132
3.8.1 模糊综合评判方法	132
3.8.2 案例	138
3.9 主成分分析	142
3.9.1 主成分分析方法	143
3.9.2 案例	147
3.10 聚类分析	150
3.11 档次数确定法	153
3.12 一点展望	158
 第4章 评估指标体系的简化	 161
4.1 指标体系简化的意义与准则	161
4.1.1 指标体系简化的意义	161
4.1.2 指标简化的准则	164
4.2 指标简化判据分析	168
4.2.1 相关性判断	168
4.2.2 效度判断	176
4.2.3 重要性判断	178
4.2.4 可信度判断	181
4.2.5 可区分度判断	185
4.2.6 可测性判断	189
4.3 指标简化的方法	190
4.3.1 指标合并	192
4.3.2 指标删除	194
4.3.3 指标添加	195
4.3.4 指标替代	196
4.4 指标简化的检验	198

4.4.1	内涵检验	199
4.4.2	保序性检验	199
4.4.3	保值性检验	200
4.4.4	共识检验	201
第 5 章	对评估者的评估分析	203
5.1	评估者的作用与影响	203
5.1.1	评估者在评估中的地位和作用	203
5.1.2	由评估者心理因素引起的各种评估误差	204
5.2	评估者的信度	206
5.2.1	评估者信度的定义	206
5.2.2	评估者信度的估计	207
5.3	评估者的选取	216
5.3.1	评估者选取的原则	216
5.3.2	评估者数量的确定	217
第 6 章	评估模型的检验、拟合与不确定性处理	221
6.1	评估模型的检验	221
6.1.1	检验	222
6.1.2	检验什么	222
6.1.3	依据什么来检验	224
6.2	拟合与不确定性处理	226
6.2.1	拟合评估模型	226
6.2.2	不确定性处理	229
参考文献		239

第 1 章 评估工作常规方法论

评估是各领域的管理决策中不可缺少的重要工作，本书对评估理论进行的探讨从评估工作的常规方法开始。所谓常规方法，是以计算加权综合分为核心的评估方法，在社会各领域的管理决策分析中被广泛运用着。本章除介绍它的具体内容外，还陈述了若干应用案例。

由于评估的理论有许多来自教育领域，也由于我们近年来大多从事高等教育评估工作，在本章中也介绍了教育评估的一些知识。然而，应该明确指出，常规方法不但被用到教育评估工作中，也被应用到许多其他的评估工作中。

1.1 评估的基本概念

1.1.1 评估的基本概念

评估，一般是指明确目标测定对象的属性，并把它变成主观效用(满足主体要求的程度)的行为，即明确价值的过程。

《宋史·戚同文传》中有“市物不评价，市人知而不欺”，这里，“评价”是还价的意思。后来，“评价”发展成为衡量人或事

物的价值的意思。当然，在实际生活中，评价、评估、评定、评鉴、估价等概念的使用比较混乱，它们被当作同义语使用的情况屡见不鲜。此外，即使在西方，理论界和实际工作者对这些术语的用法也并不规范，在英文中就有 evaluation、assessment、appraisal 和 measurement 等词。在系统科学和经济科学中，中国学者常把“评估”与“评价”当成为内涵相同、不过运用起来略有不同的两个术语。一般来说，在经济、财政、金融等实务中，采用“评估”，既为动词，又为名词；而在系统分析、学术研究中，多用“评价”，但多为名词，而较少用作动词。

首先，不论是评价还是评估，都是对对象的价值进行评判，从字面上看，评价就是评判价值的缩略语，而评估则似乎于判定之外有估计之意。一般认为，判定是确定性较强的，而估计则是确定性较弱的。事实上未必如此。将价值评判用于广泛的社会领域，价值的定义必然是广泛的，判定不可能是很确定的，必定有较强的估计性质。因此，评价与评估在确定性程度上并没有什么原则性的区别。但是，从系统科学或运筹学的角度来看，评价与评估在运用中还是有细微差别的。评价常与理论探讨匹配，例如评价理论、评价学说、评价函数、高等教育评价等；而评估则常与实务结合，例如质量评估、论文评估、项目评估、无形资产评估等。当然，也有混用的情形，大都是在实务中用“评价”一词，例如质量评价、经济效益评价、社会效果评价等。理论方面用“评估”的情况少见，特别是方法论中的评价指标、评价指标体系、评价公式，较少被称为评估指标、评估指标体系、评估公式，但评价方法却有被称为评估方法的。不过，本书中把“评价”一律称为“评估”，以符合国内评估界人士的习惯。

评估，作为动词而言，其意义是评判价值；作为名词而言，其意义则是对价值进行评判的结论。

被评判其价值的对象称为评估对象，它可能是人、人群、机构、硬件或软件形式的产品、精神文明产品等。

对评估对象的价值进行判断并给出评估结论者称为评估者，它可能是人、人群、有某种智能的客体等。

从系统科学的角度来看，评估是一种系统工程，它的基本特征和内容可用如图 1.1.1 所示的框图来概括。

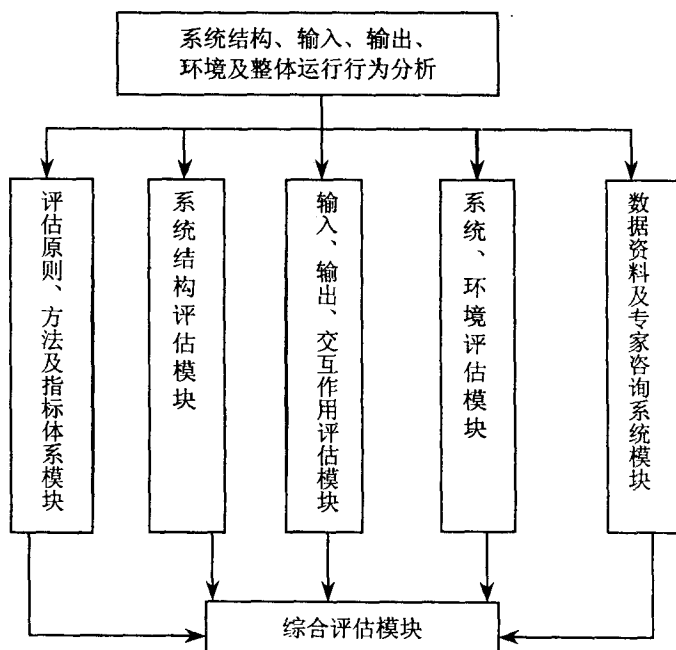


图 1.1.1 评估系统框图

当要进行某种评估时，首先要把所涉及的问题、过程、部门或体系等看成一个系统，研究其结构、输入、输出、环境以及环境与结构的交互作用、整体运行等方面，接着通过分析与构造，建立以下功能性的子系统：

- (1) 评估原则、方法及指标体系组成的工具系统模块。
- (2) 系统结构评估模块。
- (3) 输入、输出、交互作用评估模块。

(4) 系统与环境评估模块.

(5) 数据资料及专家咨询系统模块. 在任何评估工程中, 对专家的咨询以及数据处理是必不可少的.

经过以上构建, 最后便得以实行综合评估. 所谓综合评估, 包括定量分析与定性评判两部分. 定量分析通常是用计算加权综合分来实现的, 而定性评判则是根据评估工程中的各种信息(包括定量分析结果), 对于评估对象的以往表现和以后应该注意改进以及可望达到的状态给予判断性的描述.

评估因其涉及的领域不同, 而被冠以社会评估、经济评估、军事评估、学术评估、教育评估等.

1.1.2 教育评估

由于教育评估在历史上被研究得比较多, 也由于我们在本书中阐述的观点很多来源于高等教育特别是学位与研究生教育的评估工作, 在下面介绍一些有关教育评估的知识.

教育评估是一门新兴的学科, 它的理论体系、指导思想、评估模式和方法技术等, 都处于不断完善、不断发展的过程中. 因此, 虽然各国教育评估专家、学者对教育评估做了常规界说和解释, 但迄今尚未形成统一的为大家所公认的科学定义.

外国学者对教育评估有以下几种看法:

(1) 泰勒(R. W. Taylor)认为, 评估就是“确定教育目标在实际中被理解到何种程度的过程”.

(2) 布鲁姆(B. S. Bloom)等人认为, “评估是一种获取和处理用以确定学生水平和教学有效性的证据的方法.”

(3) 克龙巴赫(L. J. Cronbach)认为, “所谓教育评估, 是指为获取教育活动的决策资料, 对参与教育活动的各个部门的状态、机能、成果等的情报进行收集、整理和提供的过程.”

(4) 斯坦福评估协作组认为, 评估是“对当时方案中发生的

事件以及方案结局的系统考查——一种导致帮助改进这个方案或其他有同样总目的的方案的考查。”

(5) 美国评估标准联合会发表的定义认为，“评估是对某些对象的价值和优缺点的系统调查。”

(6) 得雷斯(Dressel)认为，“所谓评估，就是决定某种活动、目的及程序的价值过程。”

(7) 尾田睿一认为，“本来，教育评估是确认学生适合于哪个阶段的学习，确认教育成果达到何种程度。但是，现在这个概念用于广义时，则包括了全部与教育活动有直接或间接关系的各种实态把握和价值判断。”

(8) 桥本重治等人认为，“评估是与教育的目标和价值有明确关系的概念，是按照教育目标和价值观对学生的学习成绩及教育计划的效果等进行测量的过程。因此，‘评估’概念的重点在于以教育目标为标准价值判断。”

我国学者关于教育评估的界说也有十几种之多，例如：

(1) 台湾李聪明认为，“教育评估是利用所有可行的评估技术衡量教育期间的一切效果。”

(2) 陶西平认为，“教育评估就是对整个教育活动及其中各个子系统的功能进行价值判断。”

我们认为，所谓教育评估，是指按照一定的价值标准，对受教育者的发展变化及构成其变化的诸种因素所进行的价值判断。

在西方教育评估理论中，评估模式说占有重要的地位。所谓模式，乃是相对固定的评估方法论或程序。可以说，它是从具体的评估实践中抽象出来的被认为是规范性的评估程序。研究评估模式，对于探讨一般性的评估理论具有启发性。重要的评估模式包括下述几种：

1. 泰勒模式

泰勒理论的基本观点集中地表现在专门为评估而设计的所谓

的“泰勒原则”之中。“泰勒原则”首次发表于 1934 年，标题为“学生成就测验编制”；另外有关于课程发展的一般原理，1945 年以“课程与教学的基本原则”为题在芝加哥大学发行，1949 年正式出版。这两个原则归纳起来，主要包括下列几个方面：

(1) 确定教育目标。

(2) 经过若干筛选过程(如哲学、心理学、教育实验的检验)后，将选定的目标用有关术语加以表述。

(3) 找出这些行为表现的条件和情景。

(4) 确定满足客观性、可靠性与有效性诸方面的测量方法。

(5) 运用这些方法检查行为变化。

(6) 根据结果对活动、方案作出判断，并说明原因。

(7) 修正方案，重复循环过程。

泰勒模式是以目标为中心的模式，它把教育方案和计划的目标用学生的特殊成就来表示，并把这一行为目标当作教育过程和教育评估的主要依据。根据这一模式，教育评估就是判断实际活动达到目标的程度。预定的目标决定了教育活动，同时，也规定了评估就是找出实际活动偏离目标的程度，从而通过信息的反馈，促使实际工作能够尽可能地逼近目标。

2. CIPP 模式

CIPP 模式是由斯蒂弗比姆 (D. L. Stufflebeam) 在 1966 年创立的。

(1) C(Context Evaluation)即前后关系评估

前后关系评估的主要目的是评定客体的综合地位，鉴定客体的不足之处，集中和整理能用来纠正其不足之处的因素，并诊断哪些方法能提高该客体的素质。它的目的是检查教育目标和教育重点是否同学生的需要相协调，评估的最后结果应能为调整教育目标和教育重点提供一个可靠的依据。其方法由一系列客体兴趣的测量及各种类型的分析组成，一般从对客体访问着手，获得该

客体的优势、弱点和存在问题的感性认识，然后对客体再作进一步的访问，鉴定客体的行为模式和背景材料、组织特殊的诊断测验、邀请专家访问并观察、开设实验进行研究等等，最后用调查结果得出调整客体行为的设想。

(2) I(Input Evaluation)即输入评估

输入评估旨在对前后关系评估作出的改进计划能否实施作现状调查，同时对实施过程中可能出现的各种障碍作出估计，从而修订改进计划，避免它的盲目实施，以避免失败和浪费时间、人力、物力。其步骤是调查客体在满足学生的需要方面已作了哪些努力，哪些是有效的，哪些是无效的，系统地调查和分析现有人力、物力及现行的政策，收集各种有竞争价值的改进设想，并分析其有效性和可能性，综合各种改进计划的精华，提出一份确实有效、可行的改进计划。

(3) P(Process Evaluation)即过程评估

过程评估的目的是：首先，向计划人员和具体操作人员提供实施改进计划过程中的反馈信息；其次，为不断修正改进计划提供信息；第三，定期评定改进计划实施的深度以及实施对象接受的程度；最后，详细描述实施过程中的各种价值以及旁观者和参与者对总体实施质量的评判。在整个实施过程中，要进行追踪评估，定期写出报告，其目的是使改进计划及时得到修正、不断充实、不断完善，使之更符合实际需要，使实施过程少走偏路、弯路，保证实施质量。

(4) P(Product Evaluation)即结果评估

结果评估的目的是测量、判断改进计划实施结果的目标到达程度；反过来，也是对改进计划是否确实弥补了客体的不足之处、满足了学生的需求进行评估。结果评估应从广度上看待实施改进计划的效果，包括计划内的和计划外的以及肯定和否定的结果，还应收集和分析所有同实施改进计划有关的各种人员对实施结果的评判；然后，分析和解释没有达到改进计划所判定的目标的原

因；最后，从整体上评定改进计划成功的方面和失败的方面。

结果评估的手段可采用选择常模、设立对照组、进行目标测验、准则参照测验或应用行为测验，也可访问旁观者和参与实施计划的人员。为了评定超出预计的结果行为，必须对不在预计中的结果，无论是肯定的还是否定的，进行深入细致的研究，必要时进行追踪研究。

结果评估的研究报告可以是不同层次的，中期报告可在实施改进计划中期提交，内容是报告计划满足预测的需要的程度；终期报告可总结达到计划制定的目标的结果以及贯彻计划的广度，在此报告中，对实施结果要有总体分析。

结果评估的基本作用是确定一个给定的改进计划的价值程度，为制定下一个计划作出方向性指导，使计划更好地为所有有关人的需要服务，使其更有效、更有价值。

CIPP 模式的 4 个评估缺一不可，而且是循环往复，永无止境的。

3. 目的游离模式

目的游离(Goal Free)模式由斯克里芬提出。他认为，实际进行的教育活动除了会产生预期效应外，还会产生各种“非预期效应”，或者叫“副效应”。他进一步认为，根据预定教育目标进行评估，往往只注意到目标规定的预期效果，非预期效果很难得到反映。因此，为了降低评估活动中的方案、计划制定者主观意图的影响，不要把方案制定者的预定的活动目的告诉评估者，以利于评估者收集全部的关于方案的成果信息。这种不受预定活动目的影响的评估是自由目的的评估。这一评估模式与泰勒模式、CIPP 模式最大的区别是，在这种评估中，作出评估结论的依据和准绳不是方案制定者预定的目标，而是活动参与者的意图。这就是从评估活动主要反映管理者、决策者的意愿转向反映“群众”的意愿，因而具有更大的“民主性”。

4. 对手模式

对手(Adversary)模式是为了揭示方案正、反两方面的长短得失,而采取准法律过程评委会审议形式的一种评估模式.它十分重视听取关于教育方案和教育活动的争议意见,尤其是反对意见,这为各方面的情况能得到充分的反映提供了保证.这一评估的基本特点是它充分反映了各类人员“多元的”价值认识,是依靠人们直觉与经验的评估.

5. 达标模式

达标模式是确定达到某个教育计划目标的程度,同时,该模式也是确定改革效果的一种系统形式.哈蒙德(Hammond)的模式就属于这类达标模式.哈蒙德认为可以从3个方面来看:教学方面、机构方面和行为方面.他认为,各个方面又有各种因素,各种因素的相互作用会产生各种因素的组合,这种组合因素在某个教育计划评估中应看作是变量.

教学方面的主要因素有:

(1) 组织.在该组织中有教师与学生,目的是进行教学,包括时间与空间的成分.

(2) 内容.由某门学科内容确定的知识结构.

(3) 方法.有助于学习过程,可以分为3种,即教学活动、各种相互作用和学习原则.

(4) 条件.空间、设备和资金.

(5) 费用.供应和保养.

机构方面的主要因素有:学生、教师、管理人员、教育专家、家庭和社会.

行为方面的主要因素有:认知领域、情感领域和心理活动领域.

各种因素又都有各自的次因素,这些次因素也能直接影响教育计划.