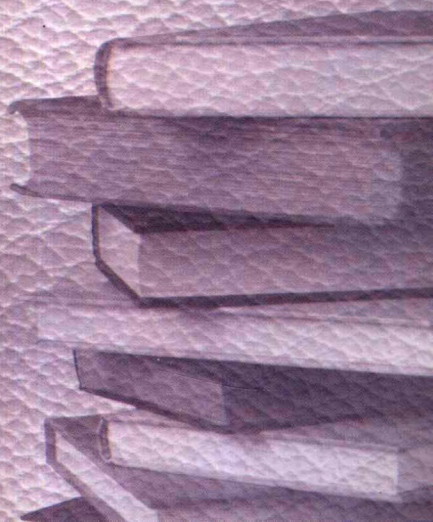


教育评估理论与方法

及在研究生教育评估中的应用

国防科技大学研究生院

国防科技大学出版社



教育评估理论与方法

及在研究生教育评估中的应用

国防科技大学研究生院

国防科技大学出版社
·长沙·

图书在版编目(CIP)数据

教育评估理论与方法及在研究生教育评估中的应用/国防科技大学研究生院. —长沙:国防科技大学出版社, 2004. 6

ISBN 7-81099-082-9

I. 教… II. 国… III. ①教育评估 ②研究生教育—教育评估 IV. ①G449 ②G643

中国版本图书馆 CIP 数据核字(2004)第 046127 号

国防科技大学出版社出版发行

电话:(0731)4572640 邮政编码:410073

E-mail: gfkdcbs@public.cs.hn.cn

责任编辑:徐飞 责任校对:黄煌

新华书店总店北京发行所经销

国防科技大学政治学院印刷厂印装

*

开本:850×1168 1/32 印张:10 字数:260 千
2004 年 6 月第 1 版第 1 次印刷 印数:1-1000 册

ISBN 7-81099-082-9/G·8

定价:18.00 元

编 审 委 员 会

主任委员：曾 淳

副主任委员：陈朝辉 王正明 王维平

委 员：苏成民 曾砥平 张春元

主 编：邵福球 曾立新 颜廷富

内 容 简 介

本书较全面、系统地阐述评估理论与方法及其在研究生教育评估中的应用。内容包括：教育评估观的演进与评估发展的世纪回眸，教育评估研究方法论，评估类型、模式和程序，指标理论与评估指标体系，评估模型与方法，指标权重理论与计算方法，评估结果的处理，评估方法在研究生教育评估中的应用。

本书不仅对从事研究生教育评估工作的人员有参考价值，对其他进行方案评估、绩效评价等领域的同志亦有裨益；也可作为高等院校教育、管理等学科专业高年级学生和研究生相关课程学习的参考教材。

前 言

教育评估工作在我国学位与研究生教育事业的发展历程中，具有无可替代的重要作用。实践表明，作为质量管理的重要手段——教育评估，发挥了国家与军队对研究生教育政策的导向作用，强化了研究生教育战线上的领导、教师和管理者的质量意识，增强了培养高素质新型军事人才、建设中国特色、世界高水平一流大学的信念和使命感。

通常，教育评估活动由四个评估要素组成：一是实施评估活动的评估主体；二是作为受评对象的评估客体；三是评估活动中所遵循的评估依据——评估标准和指标体系及其权重集；四是评估活动中所采用的模型、方法、技术和手段。评估标准和指标体系是联系主体与客体的桥梁，一方面要反映评估客体的本质属性，另一方面又要体现主体对被评对象的需求，它是衡量被评对象的准则，是评估主体进行价值判断的依据，也是被评对象在研究生培养过程中的行动准则和指南。选好评估专家，明确评估对象，制定好评估标准和指标体系，采用科学的方法、先进的技术和手段，是完成评估工作不可缺少的四个要素。

国内一些教育评估专家、学者，已为我国学位与研

究生教育评估理论与评估技术的研究、发展和应用做出了开拓性的贡献。面对研究生扩招和培养质量上经常性的监控与激励，今后的评估工作将日益繁重，更要求适用于多类型评估的理论和方法。军队学位与研究生教育中心，承担军队学位委员会授权开展的学位与研究生教育质量检查、评估，以及军队优秀研究生论文评选等工作的组织实施，也需要理论与方法上的参考与借鉴。

本书旨在移植、探索、开发新的评估模型、方法和技术，为不断改进、完善研究生教育质量评估做些努力，并力求在方法、技术上有所提升。

本书是在军队学位委员会办公室的鼓励、支持、组织与领导下编写的；国防科技大学研究生院的全体同仁鼎力协助，对本书的编写提出了许多宝贵的意见，并提供了大量信息资料；军队学位与研究生教育研究中心的同事进行了认真的阅校，在此一并表示衷心的感谢。另外，本书在编写过程中援引、借鉴和综述了一些专家学者的思想和观点以及他们的研究成果，对此深表诚挚的谢意，注释和参考文献中恐有疏漏之处，亦希见谅。感谢国防科技大学出版社对本书的出版所给予的大力支持与帮助。

曾淳

2004年6月

目 录

前 言

第一章 教育评估观的演进与教育评估发展的世纪回眸

- § 1.1 西方教育评估观的演进 (2)
- § 1.2 中国教育评价发展 (15)
- § 1.3 军队学位与研究生教育评估工作梗概 (35)

第二章 教育评估研究方法论

- § 2.1 教育评估的基本概念 (44)
- § 2.2 评价行为与评价范式 (49)
- § 2.3 综合集成方法论 (51)
- § 2.4 评估的一般理论 (54)
- § 2.5 评估的原则 (58)
- § 2.6 研究生培养工作的绩效问题 (62)

第三章 教育评估的科学性

- § 3.1 教育评估科学化的诸要素 (65)
- § 3.2 研究生教育评估的外部保障 (70)
- § 3.3 评估科学化的检验——元评估 (73)
- § 3.4 评估专家的选择是科学化评估的重要保证 (83)

第四章 评估类型、模式和程序

- § 4.1 评估的类型 (87)
- § 4.2 评估模式 (93)
- § 4.3 评估的程序 (97)

第五章 指标理论与评估指标体系

- § 5.1 指标的基本概念 (103)
- § 5.2 评估标准与评估指标体系 (107)
- § 5.3 评估指标体系的建立 (111)
- § 5.4 研究生培养的“投入产出”分析 (116)
- § 5.5 评估指标体系建立的几个参考模式 (120)

第六章 评估模型与方法

- § 6.1 集结个体成员意见的两种群决策方法 (154)
- § 6.2 群体决策支持的偏好综合技术 (158)
- § 6.3 层次分析方法 (166)
- § 6.4 综合权值排序法 (170)
- § 6.5 模糊等级区间综合评估法 (174)
- § 6.6 决策被评客体的比率评估法 (177)
- § 6.7 模糊综合评判法 (179)
- § 6.8 有限个被评对象的个体排序方法 (188)
- § 6.9 基于估计被评客体相对位置的排序方法 (194)
- § 6.10 基于图形表示的评估法 (199)

第七章 指标权重理论与计算方法

- § 7.1 权重的有关理论 (201)

§ 7.2	个体专家评估法	(209)
§ 7.3	群体专家综合法	(212)
§ 7.4	特征向量法	(216)
§ 7.5	逆称方阵法	(218)
§ 7.6	权的最小平方法	(219)
§ 7.7	区间统计法	(222)
§ 7.8	熵权	(225)
§ 7.9	指标权重体系	(230)
第八章 评估数据与结果的处理		
§ 8.1	结果的均值加权求和	(234)
§ 8.2	权威系数的确定	(235)
§ 8.3	基于专家评估准确度的评估数据处理	(238)
§ 8.4	评估主体意见的趋同原则	(240)
§ 8.5	评估结果的统一处理	(243)
§ 8.6	评估结果出现的所谓逆序现象	(250)
§ 8.7	评估数据的统一处理	(252)
第九章 评估方法在研究生教育评估中的应用		
§ 9.1	TOPSIS 法在选优评估中的应用	(255)
§ 9.2	“多数占主”权重法在教学质量评估中的应用 ..	(259)
§ 9.3	层次分析法在研究生课程质量评估中的应用 ..	(267)
§ 9.4	一级学科学位授权单位的自我评估	(275)
§ 9.5	学术声誉调查与评估	(282)
§ 9.6	博士学位论文通信评审结果的后处理	(283)

§ 9.7	博士生创新能力评估	(287)
§ 9.8	模糊综合评判在博士学位论文评阅中的应用 ...	(289)
§ 9.9	二级模糊综合评判在研究生课程教学质量评估 中的应用	(290)
§ 9.10	基于估计相对位置的排序法应用引例	(295)
§ 9.11	一级学科整体水平评估	(300)
参考文献		(305)

第一章

教育评估观的演进与教育评估发展的世纪回眸

教育评估是一门古老而又年轻的新兴学科。在中国，其渊源可追溯至隋朝开始的科举考试制度。从“五四”运动前后到 20 世纪 30 年代，是我国教育测验运动的高潮。新中国成立以后，教育上也是“一边倒”，全盘学习前苏联，一定程度上桎梏了教育评估理论的发展。直至 1977 年，教育评估理论研究工作才开始正式启动。教育评估作为一门独立的分支学科，主要始创于 19 世纪末和 20 世纪的西方。教育评价理论体系从简单到完善，评价模式从单一到多元，发生了一系列嬗变。

本章主要介绍西方教育评估观的演进、我国教育评估特别是学位与研究生教育评估活动的沿革（综述和援引袁韶莹、王琰春和陈玉琨等专家学者的研究成果），以及我军研究生教育评估活动概况，从中得到启示，以推进我军学位与研究生教育评估工作水平向更高层次发展。

§ 1.1 西方教育评估观的演进

一、教育评价的几个阶段

(一) 教育评价萌芽阶段

在 19 世纪下半叶和 20 世纪最初三十年,教育工作者对学力测验的研究和教育成就定量化、客观化、标准化方面取得了很大成绩。在回顾美国高等教育评价发展史时,首先应当提到的是,美国教育家霍勒斯·曼(Horace Mann)于 1837 年提出向英国伦敦大学学习,改变过去的单纯用口试评价学生成绩的做法,他主张以具有同一客观标准的笔试法取代带有教师主观色彩的口试法。

(二) 教育评价测定(测量)阶段

1904 年,著名教育家桑代克(E. L. Thorndike)在《心理与社会测定导论》一书中提出了著名论断:“凡是存在的事物都有数量,凡是有数量的事物都可以测量。”此论断对教育测定有很大推动作用,桑代克被后人誉为“教育测定之父”,成为美国教育评价领域的先驱者之一,他主张教育测定必须做到客观化及标准化。在桑代克的倡导下,教育测定运动在美国迅速发展起来,美国全国教育研究学会(NSSE)开始把心理学的测定应用到对教育结果的测定中去。

1923 年,心理学家凯利(T. Kelly)设计出“斯坦福大学成绩测定标准”,这是首次出现的可以综合应用于全部课程的测定标准,人们给予很高评价。1924 年,美国六个地区都成立了

“学院和中学联合会”，开展高等学校的认可鉴定工作。这实际是把教育测定的各种手段应用于对高等学校的教育质量进行鉴定和评价。这时，美国的教育测定运动达到了高潮。

到了 20 世纪末期，教育测定的弱点就逐渐暴露出来：教育测定可以评价学生的成绩与学力，也可以对教育成就进行量化、客观化、标准化的评价，却无法测定学生的世界观与思想感情等等。

（三）教育评价的描述阶段

这一时期评价的特征是对“测定”结果进行“描述”，并力求教育结果标准化。1942 年，泰勒（R.W.Taylor）与史密斯（E.Smith）合著的关于八年（1933—1941）研究的《评价委员会报告书》被人们称之为“教育评价宣言”，因为这份报告书正式提出了“教育评价”的概念，认为教育评价就是衡量实际教育活动达到教育目标的程度，而测定只是它的手段。报告书中提出了五个观点：（1）教育是使人的行动方式发生变化、得到改善的过程；（2）各种行动方式的变化都是教育的目标；（3）教育计划要以教育目标实际完成的水平而加以评价；（4）人的行为是复杂的，既不能用个别的名词概念来说明，也不能用单纯的测验来判断，对它的评价必须从多方面加以综合才能完成；（5）要搞好评价，单靠纸和笔是不够的，还应当采用观察、谈话等多种方法。

“八年研究”适应了美国社会发展的需要，获得了很大的成功。人们把这一时期称之为教育评价史上的“泰勒时期”，把他提出的关于教育评价的理论称之为“泰勒模式”，又称之为“行为目标模式”。这是一个以目标为中心的模式，就是说，预定的目标决定了教育活动，同时也规定了教育评价就是判断实际教育活动达到或偏离目标的程度，从而通过信息的反馈来促进实际工作，使之能够尽量地靠近目标。这是一种与系统分析法最为接近

的理论，它成为 20 世纪 30 ~ 50 年代前期占统治地位的教育评价理论。

1948 年，在波士顿成立了“美国心理学会总会”，学者们开始研究建立教育目标体系。1956 年，著名心理学家、芝加哥大学教授布鲁姆（B.S.Bloom）发表了《教育目标分类学（一）》。他提出，教育目标的第一级分类应当包括三个领域：（1）认知领域（包括识记、理解、应用、分析、综合、评价等）；（2）情意领域（包括注意、反应、价值倾向、组织等）；（3）精神运动领域（包括身体运动、协调动作、言语行为、非语言交流等）。布鲁姆的教育目标分类学是教育评价的重要理论基础，它对于完善教育评价理论、有效地开展教育评价工作，起了很重要的作用。

（四）现代教育评价阶段

进入了 20 世纪 60 年代，人们开始对泰勒的评价思想进行反思，特别是 1957 年前苏联人造卫星上天引起了美国的震动。美国各界人士一致认为国家科技的落后应归咎于教育、特别是高等教育的落后，因此教育必须改革。一些学者重新审视教育评价理论，认为“泰勒模式”有个根本缺陷，即教育评价是以目标为中心，而目标的合理性与可行性却是无法衡量的。何况，每项教育活动除了要达到预期目标之外，还会产生各种非预期的效应与效果，对它们也应进行评价。应以把握教育现状为主要目的，评价的领域应当包括教育目标、教育过程、教学组织、教育预算等等。这些主张对美国教育界的影响颇深。下面即介绍几个有代表性的评价观。

1. 克龙巴赫的评价观

泰勒模式在历史上占据了很长一段时期的统治地位之后，开始遭致众多的批评和反对。因为仅以目标为评价的出发点和最终归宿，似乎有悖于教育评价的初衷。为使人们注意到评价的全部

作用，克龙巴赫（L.J.Cronbach）把评价广义地定义为：“为作出关于教育方案的决策，收集和使用信息。”他认为评价并不是只调查某一教程有效还是无效，而是要确定教程需要改进的方面。

根据克龙巴赫的观点，评价的重点应该放在教育过程之中，对教育决策给予必要的改进，而不是只关心教育过程结束之后目标到达的程度。克龙巴赫的观点在教育评价界产生了广泛的影响。

2. 斯塔弗尔比姆的评价观

1966年，教育家斯塔弗尔比姆（D.L.Stufflebeam）同样对泰勒评价模式提出异议，他的观点和克龙巴赫有着很大的一致性。他认为“教育者需要一个较广义的评价定义，而不只是局限于确定目标是否达到”，评价应“有助于更好地执行和改进我们的方案”。作为替代泰勒评价的定义，斯塔弗尔比姆将评价定义为：“为决策提供有用信息的过程”。他强调：“评价最重要的意图不是为了证明（prove），而是为了改进（improve）。”这一看法不久就受到了教育界的广泛支持。1985年，他进一步提出：“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。”

与这些观点相对应，斯塔弗尔比姆提出了CIPP模式，它把评价看作是一种工具，对教育活动的背景、信息的输入、活动过程及结果给予全面的解释、评议和估计，使方案更有效地为方案使用者服务。

3. 斯克里文评价观

1967年，教育家斯克里文（M.Scriven）批评了泰勒模式，并提出了具有更大客观性的“目标游离模式”。他认为，进行教育评价时，要考察教育方案、计划、活动的真实效果，而不是预

期效果。他主张，在形成性评价阶段，应当由教育计划、方案的执行者充当评价人；而将总结性评价的教育活动作为评价对象，把搜集资料作为评价活动的重点，还应当强调评价人的客观性，一定要客观地作出价值判断。总之，作出评价结论的依据不是教育方案制定表的预定目标，而是教育活动参与者的意图。

4. 桑德与埃贝尔的评价观

进入 20 世纪 70 年代以后，美国学者对教育评价理论的研究更加深入。许多学者对相对评价进行了批判，要求废除现行的考试、测验制度。他们提出了相当尖锐的问题：教育评价的本质是什么？它应当起到什么作用？它在现实中应当完成什么任务？

1971 年，教育家桑德（O.Thornd）提出，今后的教育评价应当向新的方向发展：过去的具有惩罚性的测验，今后要成为引导、刺激学生继续发展的评价；过去以书面测验为基础的评价，今后要发展为以观察法为重点的多种多样方法的评价；过去以考察记忆力为中心的评价，今后要发展为以创造性和问题意识为中心的评价；过去的单纯在课程结束时进行考试的评价，今后要成为与课程同时进行的合作性的评价；过去只着眼于狭窄范围的测定能力的评价，今后要变成考察认识、情意及运动技能的全面的综合评价；过去只由教师进行的评价，今后应当提倡学生进行自我评价；过去只由大学进行的人学资格审查评价，今后应当由大学与高中协作，寻求更稳妥的评价方法。

1972 年，教育家埃贝尔在专著《教育评价理论》中指出，当前存在着两种截然不同的教育评价理论，一种是“传统见解”，另一种是“进步见解”。他认为，进步的教育评价理论特点是：教育评价的目的是诊断学生成绩，促进学生学习；教育的目标是培养有创新精神的开拓性人才；学生应当有敢闯精神、有独立生活的能力、积极参加各种活动；学生的学习目标应当由学生自己来确定；教师的作用应当是唤起学生的学习热情，促进学生健康