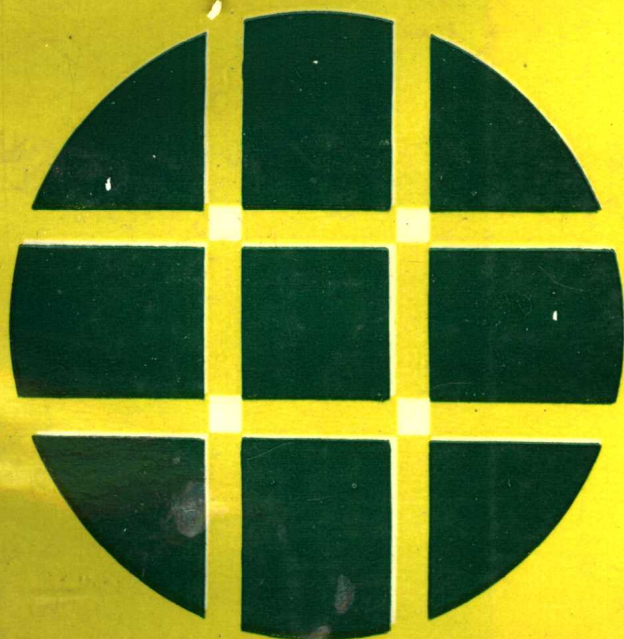


● 张 大 经
● 潘 琪
● 徐 广 文

著

高等教育评估研究与实践



中 山 大 学 出 版 社

高等教育评估研究与实践

张大经
潘 琪 著
徐广文

中山大学出版社

高等教育评估研究与实践

张大经 潘 琪 徐广文 著

中山大学出版社出版发行
广东省新华书店经销
广州轻工业学校印刷厂印刷

787×1092毫米 32开本 5.25印张 1插页 11.4万字

1989年7月第1版 1989年7月第1次印刷

印数：1—1000册

ISBN 7-306-00221-X

G·51 定价：2.95元

前 言

根据《中共中央关于教育体制改革的决定》中对高等学校进行评估的精神，1985年国家教委统一布署了高等教育评估的研究和试点工作。初期的评估试点实测工作主要在高等工程学校开展，后来其他类型的高等院校也进行了大量评估研究和实践，特别是课程或教师教学质量评估，几乎每所高校都在进行。大家在教育评估的实践中，有了自己的体会，加深了对评估的理解，积累了宝贵的经验。

近年来，我国陆续出版了几本有关教育评估的专著，总结我国教育评估的经验，推动了教育评估工作的进展。在已经面世的教育评估论著中，影响最大的是华东师范大学陈玉琨同志编著的《教育评估的理论与技术》（广东高等教育出版社，1987），该书比较全面地论述了教育评估的理论与方法，既介绍了西方教育评估的历史与现代流派，又注意探讨我国建立教育评估制度应有的特色。由北京市高教局等几个单位合编的《教育评估的理论与实践》（北京航空学院出版社，1987），是从《评估高等学校工作状态理论与方法学术讨论会》论文中，挑选、整理、编纂而成。该书虽不是一本完整、系统的论著，但也有自己的特色，具有鲜明的实践性和实用性。

本书收集了我们近年来有关教育评估的研究成果，部分文章已在书刊上发表过，也可以说是一本论文集。过去写的

文章，在编纂过程中只作少量的增删，基本上保持原有的风貌。随着教育评估研究的深入，有的问题已有了新的认识。本书是我们近四五年进行教育研究的记录和极初步的探讨。我们是在业余时间进行教育研究的，教育理论水平和实践范围都十分有限，浅薄与谬误在所难免，恳请专家和广大读者不吝指教。

本书由张大经、潘琪、徐广文编撰，苏一凡同志也参加了部分工作。在教育评估研究过程中，我们得到北京航空航天大学、华中理工大学、浙江大学等兄弟院校高教研究单位和有关同志的支持与帮助，得到我校高教研究室和各级领导的关怀与指导，在此表示衷心的感谢。大家都知道，搞教育评估研究，不能闭门造车，也不只是在名词、术语上做文章，还需要进行大量的实践。经验表明，没有领导和同行的支持与协作，评估工作是很难顺利进行和不断深入的。在课程评估研究、评估体系研究、社会评估研究和评估数据收集等工作中，还得到我校广大师生，特别是我校化学系师生和校友的支持与配合，借此机会特致以由衷的感谢。

我国的教育评估虽已进入试点总结阶段，但是看来还要大力加强评估的宣传与普及工作，还要使更多的同志了解、支持、参与评估。要建立具有中国特色的教育评估制度，不能照搬、不能等待，只能靠我们大家不懈的努力和不断的实践。

著 者

1989年春于中山大学

目 录

前 言

第一章 教育评估理论研究	(1)
第一节 高校评估的原则.....	(1)
第二节 教育评估数据的采集.....	(10)
第三节 评估量化中的若干问题的研究.....	(20)
第四节 评估的动力与政策.....	(28)
第五节 教育评估促进教育质量的提高.....	(36)
第二章 不同层次的评估	(45)
第一节 高校工作水平的评估.....	(45)
第二节 学系评估指标体系方案的研究.....	(54)
第三节 课程评估表格的设计.....	(65)
第四节 听课与课程评估.....	(74)
第五节 实验课程的评估.....	(79)
第六节 主讲教师教学质量的评估.....	(89)
第三章 若干单项评估	(99)
第一节 高校德育评估.....	(99)
第二节 社会评估的作用与方式.....	(108)
第三节 普通高校体育评估.....	(117)
第四节 高校图书资料工作的评估.....	(125)
第五节 高校教材的评估.....	(134)
第六节 教师工作的考核及评估.....	(145)
第七节 外籍教师的管理及评估.....	(152)

第一章 教育评估理论研究

第一节 高校评估的原则

确定评估体系的指标，需要遵循一定的原则。对这些原则认真讨论，有利于建立科学的、符合我国国情的评估体系。

北航的岳家俊同志认为，确定评估高等学校工作状态指标体系的原则有：(1)主要性原则，(2)可度量性原则，(3)可比较性原则，(4)可信任原则，(5)指向性原则，(6)可组合原则，(7)规范化原则，(8)外向效果原则。华东师大的陈玉琨同志对指标系统的设计原则作了比较详细的论述，特别是指标的设计应与目标的一致性，系统内指标的相互独立性问题，作了特别清晰的叙述。还有许多同志对评估的原则从不同的角度，提出自己的见解。

我们认为，作为“原则”本身应该简明、概括、权威。有些原则可以包含在另一个原则之中，同时，各“原则”之间也存在紧密的联系，例如，能否用“科学性”来概括“可信任原则”与“主要性原则”。既然是“原则”，应是不同方案，不同指标应该共同遵循的。在目前，我国教育界与学术界对如何评估尚有不同的看法，有的主张应进行“目标评估”，有的主张应将“目标评估”与“过程评估”、“条件评估”

结合起来加以考虑。“外向效果性原则”似乎侧重于“目标评估”。因此，评估原则可归纳为：科学性、方向性、可行性，基本上可概括评估的主要原则，现试分别加以论述。

一、科学性原则

指标体系应具有“整体完备性”，应该包括学校主要工作的每一重要指标。高校是一个多层次、多目标、多因素的复杂系统，缺乏足够的指标不能正确、全面地评价高校的实际水平。例如，以师资力量来说，如果只有“名教授”的有关指标，而不考虑是否“形成梯队”，不看一般教师的水平，就不容易反映教师队伍的真正水平。又如实验设备，即使有不少高级精密的进口仪器，但基础课的光度分析实验，连古老的72型分光计，也要七八位同学共用一台，实在谈不上优良的实验设备。

具有科学性的指标，应符合“主要性”原则，既要“整体完备”，也要“避免繁琐”，应该忽略那些次要的、重复的指标。有人提出某指标的权重小于其他指标的一个数量级以上，则此指标可以忽略。这种提法有一定的道理。因为权重过小的指标，就属于“次要的”指标。以经综合的“高等工业学校办学水平评估指标体系”的草案（简称《草案》）为例，多数的第三级指标的权重是0.05或0.10，对于那些权重只有0.01、0.02的指标是否可以忽略。例如“省市以上体育比赛获奖”这一指标，权重只有0.01。对学生的“体育”指标，除体育院校外，一般适宜以“身体素质”为主要指标。许多同志已经指出，有的高校在招生时为了抢“体育尖子”，简直到了不择手段的地步。如果一个理工科学生，主科的成绩不是“不及格”就是“勉强合格”，即使能获得数项体育

冠军，也实在不应认为是“人才培养”的成绩。所以删去此项指标可能更加合适。又如，“在校学生与国家计划应达的学生数”（权重为0.02），这个指标，可以包含在“师生比”、“人才培养费用”等指标之内。

科学的指标，不应存在互相重叠或互为因果的指标，例如《草案》中，由“博士（硕士）学位研究生的导师人数”，可能导出高的“博士、硕士点与学校学科数之比”。又如，一个学部委员，可能又是教材编审委员，又是学会的理事，又是研究生的导师，往往身兼数职。如果几个指标都得到体现，就可能夸大其权重。

科学的指标，应具有“优化”的结构，能够用“最少数量的指标，最大限度满足评估的需要”。但是到底多少指标才合适呢，这需要进行试验研究。我们曾对“主讲教师教学质量的评价”栏目（指标）的多少进行过研究（柳大南，中大《高等教育研究》1984、2）。栏目的基本内容相同，A表有33个栏目，B表只有15个栏目，对同一个教师进行测试，统计结果说明，测定存在系统误差。用F检验法，以 $\alpha = 0.05$ 计，两组数据有显著性差异，但若以 $\alpha = 0.01$ 计，则无显著性差异。换句话说，若要求略低时，只用一半的指标，也可以得到大体相同的结论。为了确定评估指标的数目，最好通过实验，例如设计71、65、60、50等不同数目的指标，对同一单位进行测试，求其相关情况。因为对全国来说，每减少5—10个指标，就意味节约成百上千小时的宝贵时间。

科学的指标，应具有可比性，分类、分项都可以加以比较，是否可以认为应包括“可比较原则”和“可组合原则”。不同类型、不同条件的学校的评比，显然是不公平的。为了

解决这个问题，有的同志将指标加以“当量化”。从研究的角度，这种方法有其价值，但是如将它作为主要的评估方法，似难以从本质上反映其实际水平。例如我国的田径水平，从运动员的数量看是很可观的，但是以奥运会的标准来衡量，差距是明显的。以学生的情况来说，1个博士生可以“当量为”3个本科生，但是大家都明白，1个学校“多招3—5个本科生”与“（多）招一名博士生”的水平会有什么区别。

科学的指标，应有明确的科学性的内涵，试以“基础研究项目数与科研项目总数之比”这个指标简单加以讨论。刘道玉同志对高校的科研问题曾作专门论述（武大《高等教育通讯》1982、3）。一般将科研分为：基础科学、技术科学和发展研究三大类。美国基础科学研究的投资只占12%。刘道玉提出“基础与应用并重的方向”，但是从他所举的武大本身的情况，基础研究24项，技术研究67项。所以“并重”并非指“各占一半”。不难看出，我国高校不能以基础研究为主。所以，“基础研究项目数与科研项目总数之比”不是越高越好，而且不同的学校也难以规定“最佳比例”，我们建议删去或修改此项指标。

科学的指标，应符合教育学与统计学的原理。教育具有延续性与后效性。调查毕业生的情况，可以有效反映人才培养的水平，但是毕业生离校的时间越长，社会的、家庭的、本身的各因素的影响逐渐增大，所以我们认为，适宜统计毕业后不久（例如2—3年）学生的情况。高校评价，涉及教育统计，如何运用统计学的原则有讨论的必要。例如有的学校在评估时，采用类似体操比赛评分的办法，简单地去掉最高

分和最低分，是否应该改用误差理论中“极端值舍去原则”加以计算呢？又如，在分项评分时，以1、2、3、4、5记分。但最后的结果则表示至小数点以后第二位，即结果多保留二位有效数字。其物理意义是“靠可疑数字决定等级（或名次）”，例如将总分为4.05分列为一类，而3.94分则列为二类。这样统计是否合适，似值商榷。

科学的指标，应从教育实践中提出，经过必要的试验，能经受实践的检验。通过定期的评比，应能促进教育质量的提高，促进我国教育事业的发展。“评价的基准必须争取教师群众的同意；评价的基准不是固定不变的，要不断检验和修正”（张天麟，北京科教所《教育研究资料》1985、1）。所以，提出方案应该充分论证，经过试验，但不应要求一次提出十全十美的方案。

二、方向性原则

评估的指标，应与教育总的目标一致，应与评估的目的相一致。评估既要吸取世界各国评估制度的长处，又要坚持社会主义方向，应为我们未来事业发展培养德智体全面发展的。人才。评估似应立足当前，顾及长远，面向世界。

在目前情况，评估指标的制定，应有利“多出人才，出好人才”，应该努力挖掘学校的潜力，争取各方面的支持，为“四化”多培养人才。“人才培养的费用”是一个可取的指标。据调查，1982年部分院校平均每个学生支出教育经费2351元，其中理工科大学每人支出最高4952元，最低2005元，相差347%（傅孙双，《高等工程教育研究》1984、1），减少教职工超编人数，改善师生比例，可以降低培养学生的成本。但将超编教师流向学校行政部门，似乎不是合理的方向。

鼓励学校多种形式办学，有利于学校为社会多出人才，也为学校提高效益，似应适当提高有关评估指标的权重。以中大而论，近年来，除招收本科生1.03万人外，还招收了委托代培本科生200多人，专科生300多人，干部专修科生400多人，夜大学生700多人，刊授生10余万人。但是各系极不平衡，将它列为指标，十分有利提高各系多种形式办学的积极性。

评估工作也应有利于“出好人才”。从1982年起，我国在四年里，大学招生数翻了一番，今后应控制增长率，招生数增长过快，非但不是我国的国力所允许，而且学校后勤设施也不能适应需要，所以今后高校还应十分注意提高所培养人才的质量。美国专家对他们高等教育未来提出“应改变目前只强调数量的倾向；高教要促进美国经济向以高级技术为基础的经济形式的转变，为其培养合格的人才；努力满足本地区的社会需要和解决面临的社会问题”等（斯托弗，《教育文摘》15期，1984、8、5）。对我国现今的高等教育，也面临类似的问题。评价人才的质量，可以从在校和毕业后两个阶段进行。所以在指标中规定“基础理论掌握程度”、“毕业论文的质量”等项目，并给予较大的权重是完全必要的。评价的困难或关键，是要有科学的、权威的、统一的标准。在校内、系内进行评估，一般比较容易；而要在校际进行教育质量的比较，难度是很大的。过去，许多学校也对自己的毕业生进行“追踪调查”，但一般比较粗糙，例如中大化学系曾对1981、1982届毕业学生的业务水平（包括理论基础、专业知识）进行调查，栏目设“较好、一般、较差”。1981届学生调查分别为77.6%，22.4%，0%。因为“业务水平”比较笼

统，等级分得过少，各单位评价标准不一致，调查结果不能期望有较高的有效性及可比性。对函授生、刊授生更应注意保证质量，防止滥发文凭。应能真正为提高整个民族的文化教育水平作出贡献。

评估工作应引导学校提高工作效率与效益。这个方面在《草案》中似未得应有重视，现举数例说明之。以“中心实验室固定资产总数及水平”为例，实际它已包括在精密设备固定资产的指标之中。近年来，许多高校（特别是重点院校）通过世界银行贷款等途径，已拥有一些大型贵重、精密的仪器，如何通过评估进一步提高这些仪器使用率，具有现实意义。特别是学校的中心实验室，先进的仪器设备最集中，评估的重点不宜是它拥有仪器的数量，而应该侧重于这些仪器如何为教学与科研服务。又如对科学研究的评估，同样也存在效率的问题，如果高校的科研，既不看所耗的成本，又不看投入的时间与人力，更不看对提高教学质量有何促进，这样的评估可能产生消极的结果，是可以预料的。

评估工作应该引导从我国宏观方面来看，使高校的各层次更加合理。目前我国中等人才比高级人才缺乏更严重，所以评估应鼓励办好和发展专科学校，而不是专科学校不看条件和需要纷纷“戴帽、升级”。

对高校的评估，如同各种形式的考试一样，“指向性”十分明显，各项指标的设立，应该引导高校各项工作走向正确方向，纠正当前我国高校普遍存在的一些不良倾向，进一步提高我国高校的水平。在评估指标的设计中，如何鼓励文理科互相渗透，如何提高选修课的数量和质量，如何克服实际存在的片面追求升学率，……等等，在此不一一细加讨

论。

三、可行性原则

任何改革方案或重大决策都应进行“可行性”的研究。

施行评估的人力问题。作为试点，评估一个专业，除本校的有关人员外，还要邀请社会上的名专家，人力似不十分突出。但是我们设想一下，比如中大，有20个学系，34个专业，如果全面施行评估，要邀请多少位各行各业的专家？而全国高校全面铺开此工作，又有多少专家为评估忙碌？这些专家一般身兼数职，时间特别宝贵，所以评估工作，从全国范围来看，动员太多专家，花费他们过多时间是不合适的。所以我们以为，适宜尽量依靠本地区的力量，而且国家教委、省高教局、各高校应有专门的机构与人员。如果每个专业都要国内的名专家参加评估实际是不可行的。邀请一部分中级技术人员参加评估是合适的。

评估的时间，许多同志认为学校的评估，四年搞一次是合适的。学校的评估与学系的评估有不同的要求，但是在各系进行评估的基础上对学校进行评估是有利的，至于单项评估时间可以机动些。但是，频繁的评估与频繁的考试一样不可取。否则可能出现“钦差大臣满天飞”、“报表成灾”、“人心惶惶”等不良效果。

评估应能被学校领导和广大教师所接受和支持。如果他们觉得负担甚至反感，则评估工作很难顺利进行。在进行教学质量测量时，有的同志觉得调查结果与他们原来的估计不同，有的同志说又是在搞什么新花样。从领导到教师都有各种各样的想法。要做到评估工作都象浙大光仪系评议试点那样，“对学校领导来说，是非常宝贵的参谋意见”，是要经

过一段时间的努力的。

为了使学生的学习质量具有校际可比性，看来还得使用统测法，考试法。这就象全国高考一样，虽然问题不少，也不完善，但是相当一段时间还是被采用。对统测，影响大、牵涉的问题多，争论很大。有些同志认为高校之间的统考，

“并不能真正说明人才培养的质量”，“极不利于鼓励教师进行教学改革”。考试要能考察学生的能力，要具备符合教育测量学要求的各种“试题库”，需要有关领导有很大的决心，需要耗费许多人力与物力。我国过去有的部门曾经组织较大规范的统考，例如1982年，卫生部组织13所医学院毕业生统考，对统考数据的分析，也得出一些有价值的反馈信息，对学校的教学工作也具有积极的指导作用。

评估指标应该具有可测性。例如要使学生的“思想政治表现”的评估是可行的，如果只规定它包括政治态度、学风、道德品质、遵纪守法还是不够的，还要有更具体的可测的标准。对学生思想政治表现的定量评价与实行学校之间的统考类似，看法很不一致，分歧较大。但是评估应“尽可能做到数量化”，只有这样，才能更好提高评估的有效性与可比性。对社会科学研究的“成果推广应用效益”进行定量评价也是一个难题，类似的问题适宜组织专门的力量研究突破。评估标准的确定，其难度不亚于评估指标的确定。没有“客观、明确、稳定”的标准，评估就带有很大的主观随意性，也降低了评估结果的权威。

评估工作要做到真正可行，首先当然要有可供试行的指标体系和评估标准，但是这毕竟只属于技术上的准备，还应该有关舆论和组织上的准备。前阶段时间，对高校评估的宣传与

论还做得不够，许多教师不知道评估是怎么回事，特别是没有引起有关领导的必要重视。改革是进步的必须，但任何改革都存在各种不同的思想状态，也难免有困难与阻力，这就要做好思想动员发动工作，做好试点工作，使之逐渐能为广大干部和师生所接受和支持。《中共中央关于教育体制改革的决定》明确提出“定期对高等学校的办学水平进行评估”，所以问题不是“搞不搞”，而是“如何搞好”。前段时间，许多工科院校对评估工作做了大量工作。评估工作无疑应该包括综合性大学、师范院校。

在我国的情况下，单纯依靠学术的力量和舆论的影响，很难做好评估工作。广东省课程协作组曾经拟一套试题，希望有关学校在期末考试时加试此试题，由于多种原因，这个计划未能实现。所以评估似应有行政的力量的支持。评估组织应该带有一定程度的“独立性”，免受其他因素的干扰。评估的初期，带有更多的学术性。评估的后期应逐步制度化，应逐渐转向立法。

第二节 教育评估数据的采集

一、数据采集的意义与分类

1. 数据采集是教育评估过程的重要步骤

从评估的过程看，评估过程大致包括：评估的理论研究与评估指标体系的制定；评估数据的采集和资料的准备；评估结果的分析、处理与立法等几个阶段。我国高校的教育评估，大体处于评估指标的制定与完善阶段，今后将逐渐扩大试点和深入研究，这就需要根据评估指标采集有关数据。科

学的评估要求量化，没有充分的、可靠的数据，对一个单位的工作进行评估就缺乏依据。数据收集越完善，评估的结果就越有价值。

2. 检验评估指标方案

由于我国教育评估起步较晚，缺乏经验，前一阶段，虽然提出许多评估指标体系，但只有少数专业进行评估试点，所以指标体系还需经过实践的检验。我们在实际采集数据过程中也发现不少评估指标的问题。可以预言，在评估的结果与分析阶段，也必定可以发现更多数据采集中存在的问题，这个问题在下文将进一步论述。

3. 为将来确定“绝对标准”做准备

目前，多数学校着重于评估指标（及内涵）的研究，对指标的赋值方法和标准的研究，相对来说，慢了一步。少数院校虽也提出评估标准，但缺乏科学的论证。正如陈玉琨同志所说，“绝对评价标准的制定比较复杂，事实上，只有它才能称得上严格意义上质变的关节点”。评估“绝对标准”的确定，有些类似“标样”中各元素含量的确定。虽然它实际上也是“相对值”，但是具有必要数量的、权威的数据，使结果带有“绝对”准确的含义。评估的“绝对标准”不能脱离我国的国情，只有全面地、系统地收集各校、系的有关数据，才有可能定出科学的评估标准。由于我国过去教育统计工作的落后，目前能够直接进行统计的数据并不多。以国家教委颁发的“高等工业学校办学水平评估指标体系（草案）”为例，共有72个指标，根据我们收集的数据，可以直接收集供统计的指标仅约三分之一，大多数指标因缺乏明确的内涵，或缺乏量化的方法，不能直接加以统计。