

上海市2003青年教师教学过程评价基金课题
上海市2005青年教师项目教学研究基金课题
上海市“十五”规划网络环境研究性学习课题
教育部“十五”规划网络与传统教育优势互补子课题

集体成果

教育过程 系统测评方法与实践

殷伯明 吴海红 叶苗璞 计新明 施东华等著

Systematic Testing Theory

珠海出版社

上海 评价基金课题

上海 研究基金课题

上海市“十五”规划网络环境研究性学习课题

教育部“十五”规划网络与传统教育优势互补子课题 集体成果

教育过程 系统测评方法与实践

殷伯明 吴海红 叶苗璘 计新明 施东华 等著

珠海出版社

内 容 提 要

原来,教育评价重结果。现在,不少人开始研究过程。教育部也发文,要求“进行多元综合评价,注重发展和变化的过程,充分发挥评价促进发展的功能”。

流行的CTT经典测评理论,在典型分布、线性相关、因素独立等假设下,用信度、难度、效度、区分度、标准差、标准分、等比折算、显著性检验等,解决了静态评价的问题,如高考上海卷与全国卷分数的等比折算。

本书,为解决教育过程发展性动态评价问题,提出了STT系统测评理论,提出了数据空穴填空、大片数据修复和生成技术,及其大量成功的实践应用案例。本书理论不深,方法不难,实践容易,连续几轮用作中小学教师培训教材,适合于中小学教师、教科研工作者、师范本科生和研究生阅读。

教育过程系统测评方法与实践

著 者 殷伯明等

责任编辑 怡 天

版式设计 明 言

出版发行 珠海出版社

电话: 2515348 邮政编码: 519001

社 址 中国珠海市香洲银桦新村47栋A座二层

印 刷 上海星海印刷厂

开 本 850×1168 1/32

印 张 10.2 字数: 282千字

版 次 2005年8月第 1 版第 1 次印刷

书 号 ISBN 7-900408-00-2

定 价 28.00元

十年面壁图破壁（自序）

原来，教育评价重结果。现在，不少人在研究过程。国家教育部也于 2002 年 12 月 30 日发出通知，要求“进行多元综合评价，注重发展和变化的过程”。

一、研究教育过程评价的难度

1. 教学过程概念

20 世纪，哲学开始了对“过程哲学”的研究，形成了“过程哲学”学派。“过程哲学，强调生成、变化，现实产生于过去，并前进到新的未来。”（文献 6，P567）

“教学过程，是指在教学实践中，完成教学大纲所规定教学任务的具体程序。”（文献 8，P167）

“时间要素，是维持教学系统连续运行的、最根本的，也是最有活力的要素。从宏观上讲，学生学习或教师教授有学年、学期、单元、课时之分；从微观上讲，师生间或师生与教学媒体间的相互作用，无时无刻不在时间的流程中消长。”（文献 9，P21）

2. 研究教育过程评价的难度

教育过程，是理念上的过程，面对的概念内涵抽象、对象信息不全、关系机制不明。显然，信息不全比信息完全，关系不明比关系明确，内涵抽象比内涵具体，研究时遇到的困难更多。但是，如因信息不全、关系不明、内涵抽象，就视为畏途，就会遮掩我们的视线，束缚我们的发展，就会有許多规律得不到发现，许多可获得的精神与物质财富被埋没。

上海市浦东新区信息科技学科课题组，很想克服保守思想与不求创新的态度，克服各种各样的困难，在教育过程评价的理论

与实践创新上，去深化认识，去开拓研究。

二、研究教育过程评价，离不开数字

1. 研究教育过程评价离不开数字

教育过程的数字，是一些与时间有关、表示系统行为的数；是一些彼此联系、有内在规律的数；是一些可补充生成、有可塑性的数；是一些包含多种因素、实际背景很强的数。对这些数及其隐藏规律是否认识，就看我们的知识建构。如果与数字对面相逢不相识，那么就将与隐含的规律擦肩而过，既痛苦又可惜。

众所周知，无论认识与否，规律都客观存在，也并非人为可臆造。研究这样的数，正是为了进一步认识隐含的教育过程规律。

2. 有关成绩、分数、考试与升学率的争论

有人“谈分色变”，认为升学率是评价学校质量最差的指标。其实，学生成绩，无论百分制还是等第制，都是学生个性、学校师资、学校管理、社区文化等因素的综合反映。深入客观地研究学生成绩，可发现很多教学的客观规律而不必“谈分色变”。中考与高考，测试内容覆盖面大，效度足够，难度适当。这样的测试，对个别考生，可因某种随机原因考试失误而没考出真实水平，但对考生群体来说，则有效反映了群体所在学校的教学质量。当前，教育测量评价学、教育统计学都把学生的学习成绩，作为研究各种教育教学功能与规律的切入点。

改革开放的总设计师邓小平，对学生成绩、考试、升学，早有精辟论述(文献 27, P34, 40, 104, 105):“选数理化好的高中毕业生入学，不照顾干部子弟。这样做，要是犯错误，我首先检讨。这不是复旧！一点外语知识、数理化知识也没有，还攀什么高峰？”“要经过严格考试，把最优秀的人集中在重点中学和大学。”“把坚定正确的政治方向放在第一位，这不仅不排斥学习科学文化，相反，政治觉悟越是高，为革命学习科学文化就越加自觉，越加刻苦。”“考试是检查学习情况和教学效果的一种重要方法，如同检验产品质量是保证工厂生产水平的必要制度一样。当

然也不能迷信考试，把它当作检查学习效果的唯一方法。”

三、经典测评理论缺少时间变量，也不适用于目标参照测验

100多年来，CTT经典测评理论（Classical Testing Theory）利用统计学各种典型分布与显著性检验，解决了考试的信度、效度、难度、区分度、标准差、标准分、常模样本、成绩分布、显著性检验等大量问题，为教育静态测评奠定了坚实的理论基础。

但是，建立在统计学基础上的经典测评理论，要求常模的样本量大、分布典型，因素与总体行为线性相关，因素与因素间相互独立。这些要求在现实中往往难以达到。例如，标准差、标准分，要求学生成绩正态分布，但在严重偏态情况下仍照用不误。而且，经典测评理论的测量指标及估计方法，大多只对常模参照测验有意义。例如，难度是定义在被试“平均答对率”基础上的，显然与常模参照测验性质吻合；而传统意义上的区分度，用于人员选拔时，就更为重要，也更充分体现了常模参照测验的特点。

20世纪60年代开始，目标参照测验越来越受到重视，然而其测量指标体系至今尚未建立，在对目标参照测验进行设计和分析时，搬用常模参照测验指标体系及估计方法，显然并不科学。

事实上，两种测验目的、性质、功用不同，指标体系及每个指标的含义和估计方法也不相同。探讨目标参照测验的概念、指标体系和估计方法，已刻不容缓。

另外，经典测评理论中没有时间变量，要在时间上预测未来发展，实在勉为其难，且精度也远远不够，因此很难用于动态过程发展性评价。1999年，浦东课题组使用CTT经典测评理论研究教育过程，也感觉到CTT理论在某些方面还有些力不从心。

德国诗人歌德指出“生命之树常青，而理论总是灰色的”。确实，自然科学中，每种理论都有产生的时代，有一定的适用范围，不能永远单靠一种理论走天下。十年面壁图破壁，已有学者在注意包含有时间变量的新兴理论。例如，文献9的第2、第3章，文献28，文献29，都为系统测评思想进入教育评价领域，

作了开拓性定性分析与早期呼唤。

四、浦东新区课题组提出 STT 系统测评思想

2000 年,浦东课题组申报教育过程评价市级课题,于 2003 年正式立项,并发表了理论、方法、教学预测、教学决策和教学调节等方面的阶段性成果,在 CTT 经典测评理论与系统论的基础上提出了 STT 系统测评思想(Systematic Testing Theory)。

内蒙古夏天、江苏王荣、山西荆卫东、宁夏祁和平等教研员;北京育才中学刘玉德老师;天津航空学院高有、天津理工大学刘承明、华东师范大学李健宁等教授;西安邮电大学陈真,浙江师范大学陶学明、郑洁和陈敏等同学(除个别老师,都没见过面)来电,来 Email 索取研究资料,要求对我们的研究作进一步了解。

本书就是课题组研究与实践的小结。书中,几十个数学模型,如高中生素质教育现状关联分析、中学生发育过程体形动态建模、北京 SARS 数学建模与未来预测等 21 项应用;动态建模中数据空穴填空、大片数据修复技术、试探法建模;离散数据导数等理论与方法;绝大多数属课题组原创,少数是课题组参与的原创。这些原创,其他书中没有,网络报道基本与浦东相关。

2001 年前,开展教育过程评价研究的,在上海只有少数老师。在全国的各种教育学术会议上,没有教育过程评价专题论文,书店里也没有相关专著。

2001 年后,教育部中小学计算机教育研究中心的《中小学计算机教育》、教育部主管的《信息技术教育》等全国性杂志,开始刊登浦东课题组有关教育过程评价的论文;中国教育学会中小学计算机教育专业委员会,中国教育学会电化教育专业委员会,教育部中小学计算机教育研究中心等组织的全国会议,采用了浦东教育过程评价论文,并多次评为全国一等奖。

2001 年,华东师范大学出版社出版了浦东课题组撰写的我国第一本教育过程评价专著《教育系统动态测评方法与实践》,从定性与定量结合上引入了系统测评思想与方法。华东师范大学

李健宁博士引用此书内容在《复旦教育论坛》发文(文献 56); 浙江师范大学计算机系陶学明、郑洁、陈敏同学组队, 引用我们发表的文章与资料, 参加 2003 全国大学生数学建模竞赛, 预测 SARS 流行规律, 获二等奖(文献 57); 天津大学研究生窦爱女完成硕士论文(文献 58), 杭州袁浦中学完成浙江省教育科学规划课题(文献 59), 江西农业大学黄建教授完成国家自然科学基金项目(文献 60)时, 都成功应用了我们的专著内容。

记得 1999 年初, 我们受上海师范大学承担的教育部“九五”重点课题邀请, 参与关联分析、变化分析和动态建模取得成功, 著名学者张民生、顾泠沅教授, 在专家鉴定意见中, 对我们的研究作了书面肯定。2003 年, 教育部中央电教馆宋成栋馆长, 在全国学术会议上介绍了浦东信息科技课题组的研究(文献 31); 周定珍研究员在教育部《信息技术教育》杂志上专门撰文(文献 32), 高度评价了我们的“北京 SARS 数学建模与未来预测”。

领导的支持, 学者的鼓励(文献 31, 32, 56, 60), 读者的来信及其参考我们的文献资料所做的成功应用与他们发表的论文(文献 57, 58, 59, 60), 激励着我们更加兢兢业业地继续前进。

五、本书的构思与追求

空谈术语, 不但抽象, 而且也不解决实际问题。另外, 对 STT 系统测评理论, 研究还刚开始, 还有一个由浅入深的过程。为此, 本书取名《教育过程系统测评方法与实践》, 不称理论只说方法与实践, 把原创的初步基础理论放在附录, 以待来日完善。

本书实践, 大多已发表并接受过检验。实践是检验真理的唯一标准, 应是教科研遵循的起码准则。本书 1、2 章大部分内容, 在常见教育测评书中能找到, 但在突出可比性难题, 指出 CTT 理论缺少时间变量、也不适用于目标参照测验, 指出评价理论非唯一和方法非唯一方面, 我们提出了观点、理由与实例。本书其他各章基本上都属于 STT 方法与实践。

本书力求通俗，讲清方法，重在实践，内容简练，便于阅读，易于应用；力求把 STT 系统测评方法与实践引进教育领域，为教育评价，增加一种有用的研究工具。

浦东新区信息科技课题组的高崇华、朱一军、张惠英、王渊博、童荣健、王浩平等老师参加了多年研究与本书撰写。十多年来，课题组获得了上海市教委教研室包长发、张文、倪东斌教研员，黄浦区顾伟业、周治仁教研员，浦东新区许国贤、金社平等众多教研员的支持与帮助。也感谢保姆杨从华女士对我 90 高龄多年卧床老母亲的长期护理。

六、走到了森林的边上

教育评价专家说“评价很重要，很专业，任重而道远。”一定程度上说明，长期沿用的 CTT 测评理论与方法，还需要做突破性研究，才能胜任复杂繁重的教育过程发展性动态评价。

CTT 测评理论需要突破，STT 测评理论的研究刚刚起步。我们与不少教研员一起走到了“教育过程动态评价”森林的边上，做了些理论、方法与实践探索，感到对浦东、对上海、对全国，都有些价值。因此，浦东课题组，力量微薄，还是自费出版了本专著。对我来说，年内即将退休，但还愿以本书奉献社会；对中青年教师而言，来日方长，即使没有经费支持，深信他们也有奉献精神，会努力研究系统测评理论与应用。

西塞山前白鹭飞，桃花流水鳊鱼肥。

青箬笠，绿蓑衣，斜风细雨不须归。

预计在 STT 系统测评理论的完善与巩固，方法的改进与提高，应用的开拓与创新，因素多、层次高、意义广的综合课题研究等方面，不久的将来一定会有新的突破。

殷伯明于浦东 2005.8

目 录

十年面壁图破壁 (自序)	1
第1章 教育测评与可比性难题	1
1.1 希腊寓言龟兔赛跑的可比性讨论	1
一、古希腊《伊索寓言》	1
二、“龟兔赛跑”引起的讨论	1
1.2 教育测评的起源与可比性研究	3
一、教育测评的起源及其测评思想研究	3
二、教育测评的早期发展与可比性研究	4
1.3 教育测评的发展	6
一、心理测量的发展	6
二、教育测量的发展	6
三、教育评价的发展	7
1.4 我国教育测评的发展	10
一、20 世纪 20 年代开始引进测验量表	10
二、20 世纪 30 年代陶行知评价乡村小学教育质量	10
三、20 世纪 80 年代开始第二个发展时期	13
1.5 近代教育测评概论	14
一、教育测量	14
二、教育评价	16
三、教育测评	17
四、教育测评相关的概念	18
五、教育评价目的及理论基础	20
六、教育评价分类	22
1.6 骡马赛跑与求解可比性难题的真实案例	26

一、可比性难题至今仍未完全解决·····	26
二、解决评价可比性难题的三个真实案例·····	27
三、解决评价可比性难题的思路和方法·····	28
第2章 CTT 经典测评理论对可比性求解·····	30
2.1 教育测量基本假设·····	30
一、心理素质的可测性假设·····	30
二、测量误差及来源·····	31
三、真分数及有关假设·····	33
2.2 测量的信度·····	35
一、信度概述·····	35
二、信度分类与估计·····	37
三、提高测量信度的方法·····	43
2.3 测量的效度·····	46
一、效度概述·····	46
二、效度分类与估计·····	48
三、提高测量效度的方法·····	55
2.4 经典测验理论求解可比性难题的方法·····	57
一、测验分数的解释·····	57
二、常模参考分数·····	59
三、结果参考分数·····	63
四、内容参考分数·····	66
2.5 解决可比性难题的试题分析技术·····	67
一、题目的难度·····	67
二、题目的区分度·····	70
三、区分度与难度的关系·····	74
第3章 STT 系统测评理论对可比性求解·····	75
3.1 系统思想与系统方法·····	75

一、系统思想方法	75
二、一般系统方法	77
三、系统工程方法	78
3.2 比较的环境与关联分析方法	79
一、关联分析的基本思想	80
二、影响空间和关联分析	80
三、关联分析四公理	81
四、数据预处理	82
五、关联系数	84
六、关联度与关联序	84
3.3 高中生素质教育关联分析 实践 1	85
一、人体是个复杂系统, 学生素质的分类非唯一	85
二、学生素质的关联分析	86
三、结论与建议	89
3.4 我国图书发行量关联分析 实践 2	90
一、我国图书发行量关联分析	90
二、对关联分析六项环境因素说明	93
3.5 浦东新区多元综合评价案例 实践 3	95
一、浦东新区计算机教育任务	96
二、多目标局势决策	97
三、浦东新区计算机教育效益的多元综合评价	98
3.6 动态过程评价与微分方程建模 实践 4	102
一、浦东新区中学生体形变化年段和性别比较	103
二、中学生发育过程体形变化动态建模	109
三、结论与建议	110
第 4 章 教育过程评价分析	112
4.1 教育过程评价的目的意义	112
一、政府关于教育评价制度改革的文件	112

二、教育系统推行过程评价的目的意义·····	113
4.2 教学过程评价的研究现状·····	114
一、国内研究刚刚起步·····	114
二、国外研究提出成长记录袋评价·····	115
4.3 教学过程评价的主要概念·····	117
一、教学过程评价的释义界定·····	118
二、教学过程评价的研究难度·····	119
三、亟待研究教学过程评价的理论、方法与实践·····	119
4.4 教学过程评价研究的内容与方法·····	120
一、研究的主要内容·····	120
二、教学过程评价的研究方法·····	124
4.5 教学过程的系统分析·····	126
一、教学过程的产生与演化分析·····	126
二、教学过程的组成要素分析·····	127
三、教学过程多信道表现与过程评价的诚信度分析·····	128
四、过程、阶段与总结性评价的系统分析·····	129
4.6 教学过程评价中量表的制定·····	130
一、教育测量工具与量表制定的历史·····	130
二、量表的制定步骤与要求·····	131
三、多信道参考量表的编制思想·····	132
4.7 教学过程评价中成长记录袋的应用·····	133
第5章 教育过程评价中的数学模型·····	134
5.1 系统建模方法分类·····	134
一、机理建模·····	135
二、数据建模·····	135
三、数据建模时矩阵的三种常用运算·····	136
5.2 教育系统常用数学模型·····	138
一、教育系统数据预处理模型·····	139

二、教育系统数据生成模型	139
三、教育系统效果测度模型	140
四、教育系统分析模型	140
五、教育系统微分方程动态模型	140
六、教育系统 Verhulst 动态发展模型	141
七、教育系统决策模型	141
八、教育系统回归分析模型	141
5.3 浦东中学生近视率现状与动态建模 实践 5	142
一、浦东新区中学生视力变化的年段和性别比较	142
二、城郊与城区、普通与重点中学学生视力的比较	144
三、中学生视力变化微分方程动态模型的建立	145
四、结论与建议	146
5.4 浦东高中教育效果 Verhulst 建模 实践 6	147
一、学校教育系统分析	147
二、浦东新区高中 Verhulst 教育效果建模	149
三、浦东新区高中教育系统运行分析	153
5.5 我国教育发展阶段建模和未来预测 实践 7	154
一、素质教育时间边界的不确定性	155
二、素质教育发展过程的阶段分析	156
三、我国素质教育发展阶段模型的建立	158
四、小 结	158
5.6 我国东西部教育区域发展建模 实践 8	159
一、我国教育发展的历史性差距	160
二、我国教育的补偿性发展和适应性发展	161
三、我国教育发展的区域模型	162
四、结论与建议	167
5.7 复杂过程的五步建模思想 实践 9	168
一、具体步骤	168
二、浦东新区中学教学效果系统预测五步建模	169

第6章 动态建模的数据生成与修复	171
6.1 数据的影响空间及其映射	171
一、影响空间与映射	171
二、不确定系统中的映射	175
6.2 动态建模中的数据空穴填空 实践 10	177
一、我国图书发行量预测优化的数据空穴填空	178
二、数据空穴填空的不同方式	184
三、不同激励的数据处理	185
6.3 动态建模中的数据生成技术	186
一、数据生成的意义和方法	186
二、累加生成	189
三、均值生成	190
四、光滑离散函数与级比生成	191
五、关联生成	195
6.4 动态建模中的大片数据修复 实践 11	197
一、北京 SARS 数学建模中的大片数据修复	197
二、大片数据修复技术	203
6.5 试探法建模的真实案例	204
一、费尔哈斯模型的来历	204
二、试探法建立费尔哈斯模型的理论依据	205
第7章 浦东数字化题库建设与教学诊断 实践 12	206
7.1 数字化题库建设的研究意义与目的	206
一、数字化题库建设的释义界定	206
二、国内外研究现状与课题意义	207
7.2 数字化题库建设的主要内容	208
一、对题库技术质量、构成的要求	208
二、题库的优化	209
三、题库建设的工作流程	210

四、命题系统结构与命题格式·····	212
五、组卷质量与教学诊断·····	213
六、数字化命题诊断系统中的数学模型·····	214
7.3 完成数字化题库建设的条件分析·····	214
一、课题组对提高浦东教育质量责任意识强烈·····	215
二、课题组人员具有丰富的教学与教科研经历·····	215
三、有领导支持与专家指导·····	215
7.4 数字化题库建设的研究步骤·····	215
一、启动阶段(2002.2—2004.2)·····	215
二、实施阶段(2004.2—2005.2)·····	215
三、总结阶段(2005.2—2006.1)·····	215
7.5 数字化题库建设的诊断思想·····	215
一、全面与基本·····	215
二、起点与发展·····	217
三、教学诊断与水桶原理·····	217
7.6 数字化题库诊断用数据库字段结构·····	218
一、诊断用题库字段结构·····	218
二、学校教学诊断数据库字段结构·····	219
7.7 对数字化题库组卷质量的诊断·····	220
一、试卷的测试目的·····	221
二、试卷必须要保证效度·····	222
三、试题的预定难度与区分度·····	223
第8章 教育过程系统调节·····	224
8.1 教育教学过程调节原理·····	224
一、教育教学行为目标的确定·····	224
二、教育教学过程反馈·····	225
三、教育教学过程调节·····	225
四、教育教学过程调节内容·····	226

五、教育教学的开环和闭环调节·····	227
8.2 课堂教学的过程测评与调节·····	228
一、课堂教学过程学生心理测量 实践 13·····	228
二、课堂教学过程调节·····	229
三、课堂教学的过程评价·····	231
四、课堂教学改进计划·····	232
8.3 班级教育的过程测评与调节·····	232
一、学校对班级的调节·····	232
二、班主任对班级的调节·····	233
三、班级教育调节的过程测评·····	235
8.4 课堂教学效果反馈调节实践 实践 14 ·····	235
一、课堂教学效果即时反馈要思考的问题·····	236
二、用问题测试模式进行目标反馈调节·····	236
三、课堂教学提问的目标反馈调节教学案例·····	237
8.5 浦东新区教学过程测评与调节 实践 15 ·····	241
一、从学生成绩看教学过程演化 ·····	241
二、从师资力量看教学现状·····	245
三、从学生情感表现看教学现状·····	247
四、从评价体系看教学现状·····	247
五、从教研室听课数据看教学现状·····	247
六、教学过程评价调节研究·····	248
七、测评与调节对策研究·····	249
8.6 浦东初中教学效果过程评价 实践 16 ·····	249
一、各初中入学学生的智商差异应该不大·····	249
二、浦东新区初中教学效果的时间动态分析·····	250
三、76 所初中教学效果的空间动态分析·····	252
四、76 所初中教学效果分布曲线·····	256
8.7 JY 中学用距均值评价班级教学 实践 17 ·····	257
一、社会对班级的评价·····	258