

教育测量与评价丛书

主编 董奇 辛涛

EDUCATIONAL EVALUATION, ASSESSMENT, AND MONITORING: A SYSTEMIC APPROACH

[荷]雅普·希尔伦斯 (Jaap Scheerens)
[荷]塞斯·格拉斯 (Cees Glas) 著
[英]萨利·M. 托马斯 (Sally M. Thomas)

边玉芳 曾平飞 王烨晖 译

教育评价与监测 ——一种系统的方法

 教育科学出版社
ESPHE Educational Science Publishing House



教育测量与评价丛书

主编 董奇 辛涛

教育评价与监测

——一种系统的方法

[荷]雅普·希尔伦斯 (Jaap Scheerens)

[荷]塞斯·格拉斯 (Cees Glas) 著

[英]萨利·M. 托马斯 (Sally M. Thomas)

边玉芳 曾平飞 王烨晖 译

教育科学出版社

·北京·

出版人 李 东
责任编辑 翁绮睿
版式设计 杨玲玲
责任校对 贾静芳
责任印制 叶小峰

图书在版编目 (CIP) 数据

教育评价与监测：一种系统的方法 / (荷) 雅普·希尔伦斯 (Jaap Scheerens), (荷) 塞斯·格拉斯 (Cees Glas), (英) 萨利·M. 托马斯 (Sally M. Thomas) 著；边玉芳，曾平飞，王烨晖译. — 北京：教育科学出版社，2017. 10

(教育测量与评价丛书 / 董奇，辛涛主编)

书名原文：Educational Evaluation, Assessment, and Monitoring: A Systemic Approach
ISBN 978-7-5191-0979-0

I. ①教… II. ①雅… ②塞… ③萨… ④边… ⑤曾…
⑥王… III. ①教育评估—研究 IV. ①G40-058. 1

中国版本图书馆 CIP 数据核字 (2017) 第 101753 号

北京市版权局著作权合同登记 图字：01-2009-1253 号

教育测量与评价丛书

教育评价与监测——一种系统的方法

JIAOYU PINGJIA YU JIANCE——YIZHONG XITONG DE FANGFA

出版发行 教育科学出版社

社 址 北京·朝阳区安慧北里安园甲 9 号

市场部电话 010-64989009

邮 编 100101

编辑部电话 010-64981252

传 真 010-64891796

网 址 <http://www.esph.com.cn>

经 销 各地新华书店

制 作 北京金奥都图文制作中心

印 刷 保定市中国画美凯印刷有限公司

开 本 185 毫米×242 毫米 16 开

版 次 2017 年 10 月第 1 版

印 张 27.75

印 次 2017 年 10 月第 1 次印刷

字 数 460 千

定 价 69.00 元

如有印装质量问题，请到所购图书销售部门联系调换。

JIAOYU PINGJIA YU JIANCE

Educational Evaluation, Assessment, and Monitoring: A Systemic Approach

By Jaap Scheerens, Cees Glas and Sally M. Thomas

Copyright © 2003 Swets & Zeitlinger B. V., Lisse, The Netherlands

Authorized translation from the English language edition published by Routledge, a member of Taylor & Francis Group. All rights reserved. 本书原版由 Taylor & Francis 出版集团旗下 Routledge 出版公司出版, 并经其授权翻译出版。版权所有, 侵权必究。

Educational Science Publishing House is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale throughout Mainland of China. No part of the publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher. 本书中文简体翻译版授权由教育科学出版社独家出版并限在中国大陆地区销售。未经出版者书面许可, 不得以任何方式复制或发行本书的任何部分。

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal. 本书封面贴有 Taylor & Francis 公司防伪标签, 无标签者不得销售。

中国基础教育质量监测协同创新中心研究成果

主 编 序

对大多数人来说，教育测量与评价是一个熟悉的陌生学科。说熟悉，是因为每个人的成长过程中都会经历无数次的教育测量与评价，小到家庭作业、随堂考试，大到会考、中考、高考，只是人们没有明确意识到这都属于一个重要的学科研究领域。事实上，该学科源于我国的文官选拔制度——科举制，大家公认这是教育测量与评价的早期形态。说陌生，是因为该学科是在西方发展并逐步成熟起来的，以1904年美国心理学家桑代克（E.Thorndike）的巨著《心理与社会测量导论》（*Introduction to the Theory of Mental and Social Measurements*）的问世为标志，历经比奈（A. Binet）、瑟斯顿（L. Thurstone）、洛德（F. Lord）、拉希（G. Rasch）、泰勒（R. Tyler）、克伦巴赫（L. Cronbach）、比贝（C. Beeby）等人的不断创新，从经典测验理论发展到项目反应理论，从强调行为达到目标的程度发展到重视价值判断与改进，从少数学者的个人研究发展到各类有相当规模的公办和民办考试与评价机构，迄今已有百余年的历史。

在许多发达国家，教育测量与评价、教育基本理论、教育发展理论并称为现代教育科学的三大研究领域，在教育政策制定、教育管理与教学实践以及各领域的人才选拔与评价过程中都发挥着重要的作用。回顾该学科的发展历史，教育评价是在对教育测量所存在问题的不断修正中发展起来的，而教育测量也在此过程中得到了更大的发展。今天，教育测量与教育评价两者相结合、各有侧重，测量是基于评价目的的测量，评价则是考虑测量结果的评价。

近年来，随着传统经济向知识经济的转型，各国政府和国际组织都越来越认识到教育和人才培养的重要性，教育测量与评价的研究与实践也日益活跃，其结果也越来越成为一些国家进行教育管理的有效手段。以经济合作与发展组织（OECD）发起的“国际学生评价项目”（PISA）为例，该项目自2000年正式实施以来，引起了各国教育界极大的关注，甚至被称为“教育界的世界杯竞赛”。一些在PISA测评中表现优异的国家，特别是芬兰，成为其他国家教育观摩的对象。芬兰的学生在

PISA 2003 的阅读、数学和科学方面的评测中均取得了第一名的成绩，从而被塑造为基础教育成功的典范。世界各国纷纷对其进行调研，分析其教育成功的原因，引为借鉴。

由于历史和现实的因素，我国在教育测量与评价领域的研究和实践非常薄弱，就学科建设来说，教育测量与评价还不是我国学科体系下的二级学科，各高校还没有建立起真正意义上的教育测量与评价专业方向，难以很好地承担培养教育测量与评价方向高级专门人才的重任。就教育评价的应用来看，我国各层面的教育改革都希望教育测量与评价能为教育决策和教育管理提供准确、丰富的信息，从而提高国家和各地区的教育水平；同时，我国又有巨大的人才鉴别和选拔需求，有 16 个部委成立了考试中心，每年有数以千万计的人参加不同类型的教育测量和评价，这些机构工作的开展急需更专业的智力支持。

因此，为满足我国教育测量与评价人才培养、学术研究与实践的需要，我们组织出版了这套“教育测量与评价丛书”，旨在较系统地介绍国内外教育测量与评价的新理论、新方法、新进展，以期丰富我国教育测量与评价理论。近期，我们希望首先较系统地介绍国际教育测量与评价的理论和方法的最新进展，使大家对这一领域的理论与技术有全面的理解。为此，我们选择翻译了由罗伯特·L.布伦南（Robert L. Brennan）教授主编的《教育测量》（第 4 版）（*Educational Measurement*）。布伦南教授是教育测量与评价领域最著名的专家之一，他所主编的《教育测量》一书中的每一章都由该研究领域非常优秀的学者撰写，反映了教育测量理论与方法的最新进展以及未来可能的发展方向。其次，我们也专门介绍了教育评价的原理、程序与方法。考虑到在我国，无论是教育管理者、研究者还是教育实践者，大家对教育评价都高度重视，但对教育评价作为一门学科领域的科学性和技术性的要求有所忽视，因此我们选择翻译了由詹姆斯·C.麦克戴维（James C. McDavid）教授等人撰写的《项目评价与绩效测量：实践入门》（*Program Evaluation and Performance Measurement: An Introduction to Practice*）和由雅普·希尔伦斯（Jaap Scheerens）教授等人撰写的《教育评价与监测——一种系统的方法》（*Educational Evaluation, Assessment, and Monitoring: A Systemic Approach*）两本书，向国内相关领域的研究人员、管理人员和实践人员系统地介绍教育评价作为一门学科的理论、技术的规范和要求。此外，我们认为，运用科学的手段对学生进行测量与评价，发现其学习中的优势与不足，有利于教师有针对性地教学，促进学生的发展与进步，因此，我们

选择翻译了由诺曼·E.格伦隆德（Norman E.Gronlund）教授等人撰写的《学业成就评测（第9版）》（*Assessment of Student Achievement*）一书，期望能对一线教师解决在教学实践中遇到的学生评价问题有所帮助。

我们希望，这套丛书能给我国教育科学的方法学带来一些启示；能促进我国教育测量与评价实践的科学化和规范化，为相关学科本科生和研究生的学习提供便利；能为众多中小学教师、校长以及教育管理工作、教育考试从业人员等提供相关的信息与指导。

目 录

Contents

前 言 / 1

第一部分 基本概念

- 第 1 章 教育监测与评价：概念、功能与背景 / 7
- 1.1 简介 / 7
 - 1.2 在教育中我们为什么需要进行监测与评价 / 8
 - 1.3 区分教育监测与评价技术类型的概念框架 / 11
 - 1.4 教育监测与评价的前提条件 / 16
 - 1.5 结论：为什么说是“系统的教育评价” / 18
- 第 2 章 教育评价的基础 / 20
- 2.1 简介 / 20
 - 2.2 评价方法学的基本问题 / 20
 - 2.3 评价理论的重要差异 / 28

- 第 3 章 | 15 种教育评价类型概述 / 35
- 3.1 简介 / 35
 - 3.2 基于学生成就测量的形式 / 36
 - 3.3 基于教育统计和行政数据的评价 / 43
 - 3.4 基于系统评审、观察和（自我）知觉的形式 / 47
 - 3.5 项目评价和教师评价 / 52

第二部分 系统监测与评价的理论基础

- 第 4 章 | 教育评价的政治和组织背景 / 57
- 4.1 简介 / 57
 - 4.2 评价政策背景的理性假设 / 57
 - 4.3 在教育改革项目中根据背景条件调整评价方法 / 61
 - 4.4 为监测与评价创造先决条件 / 71
 - 4.5 结论：匹配评价方法与改革项目特征，为系统的监测与评价创造先决条件、选择整体战略 / 74

- 第 5 章 | 评价：一种在学校层面规划和管理的工具 / 77
- 5.1 简介 / 77
 - 5.2 理性范式的再思考 / 78
 - 5.3 组织结构维度 / 85
 - 5.4 结论：学习和适应学校组织中内外部学校自我评价的中心地位 / 94

第三部分 学业成就评估

- 第 6 章 | 教育测量的基本要素 / 97

- 6.1 简介 / 97
- 6.2 测验目的 / 99
- 6.3 评估的质量标准 / 100
- 6.4 测验双向细目表 / 102
- 6.5 测验形式 / 105
- 6.6 测验和项目分析 / 113
- 6.7 评估系统 / 118

第 7 章 评估与评价中的测量模型 / 126

- 7.1 简介 / 126
- 7.2 二级计分项目的单维模型 / 127
- 7.3 多级计分项目模型 / 161
- 7.4 多维模型 / 171
- 7.5 多水平项目反应理论模型 / 175

第 8 章 评估模型的应用 / 178

- 8.1 评估中的测验等值和链接 / 178
- 8.2 多群组 IRT / 194

第四部分 对教育系统效果的监控

第 9 章 教育指标的框架——系统与学校层面 / 203

- 9.1 简介 / 203
- 9.2 分类 / 204
- 9.3 评价背景、合成水平与时间维度：对教育指标的进一步概念化 / 208
- 9.4 教育过程指标的作用 / 210
- 9.5 对教育指标的初步回顾 / 211

第 10 章	学校效能的多个视角 / 216
10.1	简介 / 216
10.2	一般性定义 / 216
10.3	效能的经济学定义 / 218
10.4	效能的组织-理论观 / 219
10.5	学校教育模式：效能获得的影响点 / 222
10.6	总结与结论 / 226
第 11 章	发达国家与发展中国家学校效能研究证据综述 / 228
11.1	简介：教育效能研究的整体设计 / 228
第 1 小节	来自工业化国家的证据 / 229
11.2	各种教育效能研究流派的结论 / 229
11.3	整合 / 238
11.4	元分析的总结 / 241
第 2 小节	来自发展中国家的证据 / 243
11.5	发展中国家的生产函数研究 / 243
11.6	对发展中国家学校效能研究的综述 / 245
11.7	面向教育规划者的学校效能模型的范围和局限 / 247
11.8	总结与结论 / 250
第 12 章	对教育起作用的影响因素的含义 / 253
12.1	简介 / 253
12.2	成就导向/高期望 / 255
12.3	教育领导力 / 256
12.4	员工之间的一致性和合作性 / 262
12.5	课程质量与学习机会 / 264
12.6	学校氛围 / 266
12.7	可评价的潜力 / 272
12.8	家长参与 / 275
12.9	班级风气 / 277

- 12.10 有效的学习时间 / 279
- 12.11 结构化的教学指导 / 281
- 12.12 自主学习 / 285
- 12.13 差异化 / 285
- 12.14 强化与反馈 / 288
- 12.15 总结与结论 / 288

第 13 章 教育增值的测量指标 / 293

- 13.1 简介 / 293
- 13.2 增值的概念 / 295
- 13.3 数据收集过程和问题 / 295
- 13.4 怎样测量增值 / 299
- 13.5 增值结果的呈现 / 304
- 13.6 将增值指标作为学校评价和自我评价的有效工具 / 306
- 13.7 实施增值评价系统：学校、地方和国家主管部门的角色 / 308
- 13.8 研究日程 / 308

第五部分 督导与学校自我评价

第 14 章 学校督导基础上的监测 / 313

- 14.1 简介 / 313
- 14.2 学校督导系统 / 314
- 14.3 学校督导的英格兰案例研究 / 317
- 14.4 结论 / 322
- 附录 为通知督导过程所收集的信息和数据 / 323

第 15 章 学校评价：基本概念 / 326

- 15.1 简介 / 326
- 15.2 定义 / 326

- 15.3 学校质量评价 / 329
- 15.4 学校评价方法的基本类型 / 335

第 16 章 学校自我评价中的问题与困境 / 340

- 16.1 简介 / 340
- 16.2 对结果的解释与应用：学校效能观点对这个问题的帮助 / 341
- 16.3 信息使用的组织与交流 / 343
- 16.4 使用的背景 / 346
- 16.5 学校自我评价结果的保密性 / 348
- 16.6 学校（自我）评价中的困境 / 354
- 16.7 执行问题：在发展中国家中的应用 / 357
- 16.8 结论 / 360

第 17 章 增值评价指标研发与使用的实践案例

——兰开夏郡教育局增值评价项目 / 362

- 17.1 简介 / 362
- 17.2 兰开夏郡教育局增值评价项目的研发 / 363
- 17.3 一所学校自评活动的案例研究 / 370
- 17.4 优秀经验总结 / 378
- 附录 A 学校增值结果表（虚构的学校） / 379
- 附录 B NCER 学科差异样例 / 381
- 附录 C 学生态度量表细目 / 382
- 附录 D 学校如何使用数据的总结 / 384
- 附录 E 案例校：学生个体监测中期报告样例 / 385

参考文献 / 388

索引 / 417

后记 / 429

前 言

XIII

在教育评价的应用中，持续的监测与各种针对学生学业成就的评价和测量比项目评价更普遍、更频繁。但是，大量的教育评价教材（textbooks）却在讨论项目评价。本书并不否认项目评价基本逻辑的重要性，包括对待评价项目的分析准备阶段、方法学在项目效果评估中的作用、结果使用的思路框架，但是本书却具有一个不同的定位。

基于一个广泛的框架

本书关注在组织、管理与教学过程中的教育评价、评估与监测的应用。本书的结构建立在一个三维模型之上，并以此为基础区分各种教育监测与评价类型。这三个维度是：评价的基本功能、在教育系统中的应用层级与数据收集策略。

我们称这个方法是“系统的”，因为它需要在多层级嵌套的教育系统中依据所采用的评价形式开展。

系统方法

“系统”更具体的含义如下：

- 系统的观点在这里指的是监测与评价在教育系统内制度化的应用，而不是仅局限于项目评价。
- 监测与评价被看作是有利于教育系统的日常运作与改进的；这种观点来自系统论的“控制论”（the cybernetic principle），控制论描述了基于评价和反馈的学习与控制。
- 监测与评价的战略性的使用取决于多层级教育系统的决策结构和权力在各层级的分布情况。
- 完整性指的是各种教育测验、监测、评价都是教育系统的必备要素，在教育系统的不同层级根据不同的目的提供反馈；监测与评价的战略性的使用是充分

挖掘各种评价形式组合的潜能之后做出的一种经济的选择。

- 教育系统的投入-生产/加工-产出模型是展示教育内容、产生教育监测与评价中关键目标领域的框架。

XIV 在“背景”中分析教育监测与评价

有两种能够将教育监测与评价活动置于 (embed in) 组织背景中进行描述的方式：理论层面的描述和实践层面的描述。在理论层面，教育监测与评价可以从治理与管理的理想模型的角度来分析，例如概要式理性规划 (synoptic rational planning)、教育市场机制的运行等等，也可以从控制论及其相关理念来分析，例如反溯式规划 (retro-active planning)、“作为学习型组织的学校”等等。这些分析框架可以更好地帮助理解教育监测与评价在治理与管理中的功能。在实践层面，可以把监测与评价发生的机构、政治、组织背景看作其限制条件。理解这些背景的局限性是克服其局限并为采用合理评价技术手段创造更佳前提条件的第一步。在国家水平，特别是在发展中国家，进行监测与评价被视为一项创新活动，往往要求进行相应的组织改革。这些改革包括招收并合理使用必要的技术人才，促进决策者与其他利益相关者使用信息。

教育监测与评价的三种核心方法学定位

简介部分 (第一部分) 概述了监测、评估与评价方法，简要介绍了系统评价的思想以及相关的基本概念。第二部分介绍了上面提及的组织条件与局限。本书在第三、第四、第五部分进一步阐述了三种主要的数据收集策略与方法：评估与测量学生成就、基于指标的学校效能监测，以及督导与学校的自我评价。

第三部分介绍教育评估与测量，对经典测验理论、项目反应理论及其在教育中的应用做了基本的介绍。

第四部分关注教育效能概念、基于各类指标体系的效能监测。

最后，第五部分着眼于学校自我评价的各种方法，对各种方法学和实际问题进行了探讨，并提供了一所英国学校自我评价的案例。这部分还用一章的篇幅介绍了当代学校督导方法。

由于本书关注系统监测的应用，关注组织背景与局限，关注与教育效能具体成果的联系，关注方法学的最新进展，所以本书适合多个读者群：它可以是教育学研