

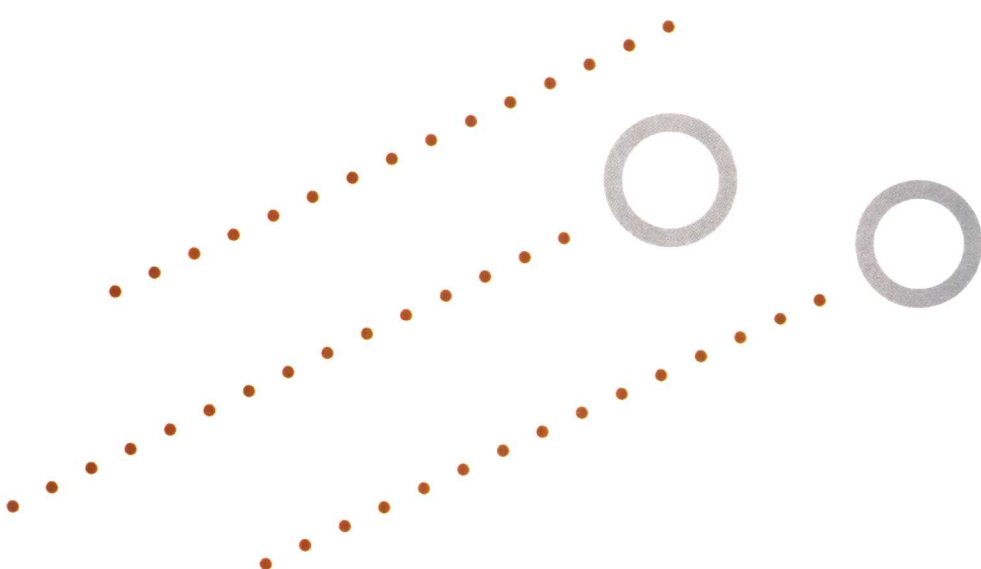
当代教师新支点丛书

Developing and Using
Classroom Assessments (Third Edition)

开发和运用课堂评估

(第三版)

【美】Albert Oosterhof 著
谭文明 罗兴娟 译



 中国轻工业出版社

当代教师新支点丛书

Developing and Using Classroom Assessments
(Third Edition)

开发和运用课堂评估

(第三版)

【美】Albert Oosterhof 著

谭文明 罗兴娟 译

 中国轻工业出版社

图书在版编目 (CIP) 数据

开发和运用课堂评估, 第三版 / (美) 奥斯特霍夫
(Oosterhof, A.) 著; 谭文明, 罗兴娟译. —北京: 中
国轻工业出版社, 2006.9

(当代教师新支点丛书)

ISBN 7-5019-5442-9

I. 开... II. ①奥... ②谭... ③罗... III. 课堂教学—
教学评议—中小学 IV. G632.421

中国版本图书馆 CIP 数据核字 (2006) 第 059445 号

总策划: 石 铁

策划编辑: 吴 红

责任编辑: 朱 玲 吴 红 责任终审: 杜文勇 封面设计: 黄金支点

版式设计: 时春雨 责任校对: 万 众 责任监印: 刘志颖

出版发行: 中国轻工业出版社 (北京东长安街 6 号, 邮编: 100740)

印 刷: 北京天竺颖华印刷厂

经 销: 各地新华书店

版 次: 2006 年 9 月第 1 版第 1 次印刷

开 本: 660 × 980 1/16 印张: 25.50

字 数: 300 千字

书 号: ISBN 7-5019-5442-9/G · 618 定价: 42.00 元

著作权合同登记 图字: 01-2005-6116

咨询电话: 010-65595090 65262933

读者服务部邮购热线电话: 010-65241695 85111729 传真: 85111730

发行电话: 010-65128898 传真: 85113293

网 址: <http://www.chlip.com.cn>

E-mail: club@chlip.com.cn

如发现图书残缺请直接与我社读者服务部 (邮购) 联系调换

51278J5X101ZYW

版权声明

Simplified Chinese Edition Copyright © 2006 by Pearson Education Asia Limited and China Light Industry Press.

Original English language title: *Developing and Using Classroom Assessments*, third edition, by Oosterhof, Albert. Copyright © 2003. All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Merrill Prentice Hall.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Region of Hong Kong and Macau).

本书封底贴有 Pearson Education（培生教育出版集团）激光防伪标签，无标签者不得销售。

译者序

课堂评估是教师教学过程中一个非常重要的环节，正如本书作者所言“没有很好的评估，我们是不可能知道有效学习是否发生了”。评估不仅能为教育行政部门的决策者提供决策信息，更重要的是为教师教学提供反馈信息。翻译整本书后，我发现本书的课堂评估完全围绕教师的教学经验而展开论述，将枯燥、深奥、难懂的测量和统计知识讲述得生动贴切、深入浅出，使读者在分享类似驾驶汽车和与朋友聊天的经历中就深刻体会到了非正式观察和提问的特征、指导原则和注意事项；也使读者在简明而清晰的示例中理解了各类考试试题的编制要求和非正式评估的编制步骤以及一些描述统计量的基本含义和用途；此外本书还特设一章讲述了学生应试的技能、技巧，不仅实用，还有利于提高学生的考试能力。像这样的书不仅适合教学决策者阅读和学习，更值得编制试卷的教研员和广大一线教师阅读和学习。尤其是新课程改革以来，教师自主教学的空间在不断扩大，课堂评估成为教师自主教学中重要的一部分，而此书正像一场及时雨为迫切需要提高课堂评估能力的教师提供了详细的指导，这也正是我们急切翻译此书的初衷。

本书的内容翔实而丰富，开篇不仅论述了评估在课堂中的作用，还分析比较了评估、评价和测量的区别以及正式评估和非正式评估的区别，这使得读者概念清晰，思路明确。在接下来的四部分内容中，该书首先概述了课堂评估框架的建立，包括评估的目标、信度、效度和评估的解释与应用。接着通过举例等各种方式详细论述了书面考试中各类常见题型的编制、实施和评分过程，具体涉及填空、选择、简答和论述题的编制和应用。如果说教师对书面考试这种正式的评估方式有一定的认识，那么对于非正式评估而言，教师恐怕还停留在感性认识的层面。本书注

意到了这一点，因此通过恰当的类比和大量的示例具体讲解了非正式评估的编制、实施和评分过程，具体内容包括非正式观察和提问以及当前较流行的表现性评估和档案袋评估等。与其他论述不同，本书在表现性评估上尤其突出表现性评估编制的三个步骤，按照这三个步骤你一定能编制出一份满意的表現性评估方案。而在档案袋评估方面，本书强调学生如何参与到档案袋评估中，这无疑为教师深刻理解和运用档案袋评估指明了方向。最后，本书的焦点集中在如何使用这些评估上，讨论了评估和教学的整合、学生表现等级的评定和汇报方式以及标准化测验等。在具体应用的例子中生动地讲解了字母等级、核对表、等级量表以及标准分、百分位数和年级当量等相关的描述统计量，使教师在认识评估方式的基础上更深刻地解释和运用评估结果。

本书的编制结构也很有特点，它特别注重读者的参与。本书每一章都针对重要的知识点设计练习题，并给出练习答案和分析过程，使你在学习的过程中可以不断地评估自己，提供反馈，促进你的学习。而每章最后，在“试一试”中又设计了许多与本章内容相关的实践活动，在实践中提高你的评估能力。总之，我们相信本书以其丰富的内容和众多的练习机会一定会使你在阅读后获益匪浅。

感谢中国轻工业出版社“万千教育”编辑部能及时地引进本书，为我们的教学评估提供了更多的途径，也为广大教师编制评估试题和改进课堂评估方式提供了借鉴。本书不仅帮助教师深刻理解课堂评估的理论背景知识，还帮助教师在实践中具体运用各种课堂评估方式，理论和实践的紧密结合极大地提高了教师的课堂评估能力。

本书的序言、目录、第一章、第三章至第十章由谭文明翻译，第二章、第十一章至第十七章以及附录由罗兴娟翻译。

由于时间紧迫，再加上我们自身翻译水平有限，书中难免有不足之处，敬请各位专家、读者不吝赐教，万分感谢。

罗兴娟

2006年春于北京师范大学

序 言

本书和相应的网站为我们提供了一些关于课堂评估开发和运用的重要且简洁的讨论。本书可以作为大学课程，但是就本书涉及的内容和所费笔墨来说，作为完全意义上的测量教科书却又不太合适。《开发和运用课堂评估》一书特别适合在以下几种情境中使用：

- 含有大量关于评估的讨论的教育心理学课程
- 评估作为其组成部分之一的综合方法课程
- 借助一系列阅读材料使学习者获得大量实践经验的全程的评估课程
- 中小学及其以上水平的教师自学时使用

本书的特点

本书期望读者参与到我们的讨论中来。为了说明这一点，请你花一点时间浏览本书，你会发现书中包含了大量练习题。每个练习题的标题都是“运用你所学的知识”（书中统一译为“练习”——译者注），并且请读者立即把刚才学习的概念运用到现实情境中。之后还会根据这些练习题进行一些讨论。本书还通过激发我们每个人作为学生或教师在评估方面具有的大量经验，让每位读者都能参与其中。书中每章的导言部分都使用我们熟悉的情境来导入即将讨论的问题，每章都以活动结束，这些活动把每章的内容与课堂的现实需要联系起来。

《开发和运用课堂评估》还综合了课堂评估发生时的大量背景知识。比如，关于认知心理学在评估中的主要应用的讨论就贯穿了全书，例如陈述性知识和程序性知识的区别、问题解决的本质和评估这些知

识的选择等等。同样，鉴于非正式评估目前在课堂评估活动中占有相当高的比例，我们特意用一整章介绍了非正式观察和提问的运用，而另外还有一章专门讲述如何把这些评估整合到教学中。

组 织 结 构

《开发和运用课堂评估》总共有四个部分。第一部分为教师提供了评估学生的评估框架。这个评估框架的成分包括确定如何使用评估结果、确立那些能表明学习是否已经发生的行为、收集证据证明评估是有效的以及确定从能观察的表现中概括出不能观察的表现是否合理。评估框架的这些成分都使用实践性的或我们熟悉的例子加以说明。评估框架在本书第二、第三和第四部分都得到了广泛的应用。

第二和第三部分涉及课堂评估的编制、实施和评分。第二部分集中讨论了书面考试，包括论述题、简答题和各种客观评分方式等。从第六章到第九章，读者将会学习何时和如何使用每一种评分方式，包括那些在编制每种试题时都不容忽视的一些重要特性。读者可以立即通过评价结构良好的和有缺点的考试题的方式来实践自己刚刚学到的知识。第十章的内容主要是帮助学生参加书面考试。

第三部分主要讲述其他评估方式，包括非正式观察和提问、表现性评估和档案袋评估。尽管这些非正式评估为课堂中的许多正式评估提供基础信息，但许多书往往对这种偶然发生的观察和提问学生都只进行粗略的介绍。而本书第十一章就详细讨论了这方面的评估，这一章讨论并确定了这种偶然发生的观察和提问的关键特征，且通过分析读者都很熟悉的一些社会性相互作用的方式举例说明了这些观察和提问的实用性。本书用两章的篇幅讲解了表现性评估。第十二章主要讲解我们在考虑何时使用这种有效的评估媒介时应该注意的几个因素，包括表现性评估的特征、评分的选择和编制表现性评估前的准备工作。第十三章详细讲述了编制表现性评估的步骤，并列举了五个表现性评估的编制示例进行说明。第十四章主要描述如何使用档案袋来整合评

估和教学，包括如何设计档案袋和如何指导学生使用档案袋。

第四部分主要讲述这些评估方式的运用。第十五章的主要内容是把评估整合到教学中，在这一章尤其要注意评估在建构主义学习方法和认知主义学习方法中扮演的具体角色。第十六章讲述了如何向学生父母或课堂外的其他人汇报学生的表现。这章讨论的问题包括从我们观察的表现推断学生知道什么和确定表现的标准以及给出分数等。第十七章主要讲解了标准化测验中学生分数的解释。这一章最重要的部分能帮助读者评价我们常用的并能直接影响课堂活动的标准化测验。

致 谢

我始终都深深地感谢在我写作本书过程中与我分享其经验的那些学生、教师和其他同事。在当前的这个版本中你将会继续发现他们的观点和见识。我非常感谢俄克拉何马州州立大学的 Laura L. B. Barnes，缅因州州立大学的 Alexander W. Chizhik 和 Montevallo 大学的 Lee Doebler，以及佐治亚州南方大学的 Marlynn Griffin。他们为本书重要部分的写作提供了深刻而又非常有用的建议。我还由衷地感谢 Merrill/Prentice Hall 全体编辑人员的慷慨帮助。我个人特别感谢 Kevin Davis，因为他的构思在本书的初版中起着重要的作用，而他的见识和指导又促使了本书第二版和第三版的出版。此外，我还感谢在本书出版过程中 Mary Harlan 提供的帮助和 Carlisle Publishers Services 提供的详细的编辑服务。

Albert Oosterhof

目 录

第一章 导言	(1)
第一节 测量的重要性	(2)
第二节 测量、评估和评价之间的区别	(3)
第三节 正式评估与非正式评估的区别	(6)
第四节 课堂评估的作用	(7)
第五节 最优表现和通常表现	(10)
小结	(11)

第一部分 如何建立评估框架

第二章 确定如何解释和运用评估	(16)
第一节 解释学生表现的参照系	(17)
第二节 标准参照解释和常模参照解释的意义	(20)
第三节 选择恰当的解释方式	(25)
小结	(28)
练习答案	(28)
试一试	(29)
参考读物	(29)
第三章 可测量的目标与目的	(31)
第一节 学习结果的分类	(32)
第二节 表现性目标的组成	(42)
第三节 表现性目标的选择	(47)
第四节 将表现性目标传递给学生	(50)
第五节 目的与表现性目标	(51)
小结	(51)

练习答案	(52)
试一试	(53)
参考读物	(53)
第四章 收集效度证据	(55)
第一节 结构关联效度的证据	(56)
第二节 内容关联效度的证据	(59)
第三节 效标关联效度的证据	(65)
第四节 有效的解释和测验的使用	(66)
小结	(72)
练习答案	(72)
试一试	(73)
参考读物	(73)
第五章 用观察到的表现概括没有观察到的表现	(74)
第一节 为什么观察不能概括化	(74)
第二节 检测不一致性的方法	(83)
第三节 提高概括能力的技术	(89)
第四节 概括能力与效度的关系	(93)
小结	(94)
练习答案	(95)
试一试	(96)
参考读物	(96)

第二部分 如何对书面考试进行编制、实施和评分

第六章 填空题和简答题	(101)
第一节 简答题的优点和不足	(102)
第二节 明确简答题要求的性质	(104)
第三节 练习应用这些性质编制简答题	(107)
小结	(110)
练习答案	(110)

试一试	(110)
参考读物	(111)
第七章 论述题	(112)
第一节 论述题的优点与不足	(113)
第二节 明确论述题要求的性质	(116)
第三节 练习应用这些性质编制论述题	(124)
第四节 为学生论述题的作答评分	(128)
小结	(131)
练习答案	(132)
试一试	(132)
参考读物	(133)
第八章 多项选择题	(134)
第一节 多项选择题的优点和不足	(136)
第二节 明确多项选择题要求的性质	(141)
第三节 练习应用这些性质编制多项选择题	(150)
第四节 多项选择题的变化类型	(155)
第五节 选项最理想的个数	(160)
小结	(161)
练习答案	(161)
试一试	(162)
参考读物	(163)
第九章 二项选择题	(164)
第一节 二项选择题的变化类型	(166)
第二节 二项选择题的优点与不足	(171)
第三节 明确二项选择题要求的性质	(175)
第四节 练习应用这些性质编制二项选择题	(180)
小结	(185)
练习答案	(185)
试一试	(187)

参考读物	(187)
第十章 学习如何参加笔试	(188)
第一节 对测验方法的熟悉程度	(188)
第二节 为测验做准备	(190)
第三节 测验智慧	(193)
小结	(202)
练习答案	(203)
试一试	(203)
参考读物	(204)

第三部分 如何对其他评估方式 进行编制、实施和评分

第十一章 非正式观察和提问	(207)
第一节 非正式观察和提问的特征	(208)
第二节 使用非正式观察的指导原则	(218)
第三节 使用非正式提问的指导原则	(225)
第四节 保留记录	(229)
小结	(231)
练习答案	(233)
试一试	(233)
参考读物	(234)
第十二章 表现性评估的适用条件	(235)
第一节 表现性评估的特征	(237)
第二节 对表现性评估的评分	(243)
第三节 编制表现性评估的准备工作	(251)
小结	(254)
试一试	(255)
参考读物	(255)
第十三章 表现性评估的编制	(257)

第一节 确定要评估的能力	(258)
第二节 确定要观察的表现	(263)
第三节 编制对学生表现进行评分的计划	(269)
第四节 编制表现性评估的具体例子	(271)
小结	(282)
练习答案	(283)
试一试	(284)
参考读物	(284)
第十四章 档案袋评估	(286)
第一节 档案袋评估的特征	(287)
第二节 学生档案袋的设计	(294)
第三节 指导学生使用档案袋	(306)
小结	(306)
试一试	(307)
参考读物	(308)

第四部分 如何运用评估

第十五章 评估和教学的整合	(313)
第一节 在帮助学生成为专家时整合评估与教学	(315)
第二节 在帮助学生建构新知识时整合评估和教学	(327)
第三节 评估和教学整合后概括化和效度的相关性	(330)
第四节 对客观评估的启示	(336)
小结	(339)
试一试	(341)
参考读物	(342)
第十六章 评定和报告学生的表现	(343)
第一节 可选择的报告方式	(343)
第二节 建立等级标准	(352)
第三节 等级的激励和约束作用	(357)

小结	(360)
练习答案	(361)
试一试	(361)
参考读物	(362)
第十七章 解释和使用标准化测验	(363)
第一节 能力倾向测验与成绩测验的区别	(364)
第二节 标准化测验分数的常用解释	(366)
第三节 在课堂上使用标准化测验	(378)
小结	(384)
练习答案	(385)
试一试	(386)
参考读物	(387)
参考文献	(388)

第一章

导 言

让我们先来计算一下自你入学读书起曾经参加过的考试数量。如果你每周在每一门主要科目上有 30 分钟的考试和测验的话，那么到你高中毕业时，你就已经参加了 1000 多个小时的笔试，这相当于连续 40 多个整天的时间。这还不算你在大学里完成的考试，也不包括从你如何握笔写字到你如何完成一项实验等等成千上万的非笔试的评估，更没有算你完成的大量的论文和计划以及你所做的笔记和档案袋。尽管学生们可能认为课堂评估对他们来说要比对教师更重要。但据估计，一位教师超过 25% 的时间都用在了各种不同形式的课堂评估的开发和运用上了。

不光是在学校的课堂环境里，其实在我们每天的日常生活中，我们总是在被评估和评估着别人。我们估摸着别人对我们说过的话，他们的面部表情、措辞以及他们的态度。我们以他人对我们的反应来评价自己的言行。我们常常研究周围有什么样的人和事，以决定我们要做什么事和去向哪里。

这样一种我们对待测验和其他类型的评估所共有的经验对于学习本书所讲述的材料是非常重要的。例如，你知道有些试卷要比另外一些试卷出得好，在我们要把这种可以编出好试卷的因素分析出来时，我们可以利用这一知识。你明白测验常常会出现一些具体的问题，例如试卷中问题设置得很模糊，以及所出的试卷的内容同课程中学习的知识不相对应。你知道，你写的作文、你所解决的数学问题以及你的艺术作品如果由不同的教师来打分数，结果也可能会有所不同。你还

知道和课堂评估的使用相关的许多问题。有些时候，教师在测验结束很久之后才公布结果；而且教师在公布结果时也没有注意将学生遇到困难的知识解释清楚，教师常常在还没有解释清楚测验中检测出的问题之前就已经匆忙地继续后面的教学了。在档案袋中的作品已经完成之后的几天，有时甚至是几周后，它们才会被教师所检阅，并且也只有其中的一小部分被仔细地检查。

我们知道教师必须且不得不依靠非正式的评估，例如不定期的观察和口头提问。在使用这种非正式的评估方面，有些教师做得比另外一些更为有效，他们对于了解将要发生的事有一些技巧。一些教师保持着对学生最初的印象，而另外一些则使用非正式的评估来检测在学生发生的变化和最初观察中的错误。一些教师有效地利用非正式的评估引导学生的学习，另一些教师却忘记了学生正在学习什么样的知识。在我们呈现编制试卷的步骤时，在我们准备表现性评估目标时，在口头提问以及观察学生行为时，我们所共有的经验是非常宝贵的。

第一节 测量的重要性

词典中罗列了动词“测量”的相关定义。对于我们的讨论最为有用的定义为：确定某种事物的属性，遵循某一标准，并同诸如他人表现一类的参照作比较。

几乎在所有的学科中，测量都是非常重要的。在物理科学中，测量用以确定不同物质的属性或是弄清一种物质何时获得某种特性。在机械工程领域，像负荷量、力矩、摩擦力以及热量性质等都是经常需要测量的。在体育运动中，测量也会用来作出决定，例如裁定棒球中的一击、橄榄球中死球的判定以及篮球中的一次个人犯规的判罚。测量也可用在政治界以预测选举的结果，或是评价另一政府对于一项政策决定的反应。在表演艺术中，测量也可以用在导演对特定角色演员的选择和评论家对于首场演出的评议上。商业中则更是要依靠测量所得的数据决定产品技术的有效程度以及对广告影响作出评价。