

教育测量与评价丛书

主编 董奇 辛涛

PEARSON

ASSESSMENT OF STUDENT ACHIEVEMENT (NINTH EDITION)

[美] 诺曼·E. 格伦隆德

C.基思·沃 著

NORMAN E. GRONLUND

C. KEITH WAUGH

教育部基础教育质量监测中心 组织翻译

杨 涛 边玉芳 译

学业成就评测

(第9版)

教育测量与评价丛书

主编 董奇 辛涛

学业成就评测

(第9版)

[美] 诺曼·E. 格伦隆德

C.基思·沃 著

NORMAN E. GRONLUND

C. KEITH WAUGH

教育科学出版社

· 北京 ·

出版人 所广一
责任编辑 周益群
版式设计 杨玲玲
责任校对 曲凤玲
责任印制 曲凤玲

图书在版编目 (CIP) 数据

学业成就评测 (第9版) / (美) 格伦隆德, (美) 沃著; 杨涛, 边玉芳译. —北京: 教育科学出版社, 2011. 12

(教育测量与评价丛书 / 董奇, 辛涛主编)
书名原文: Assessment of Student Achievement
ISBN 978 - 7 - 5041 - 5971 - 7

I. ①学… II. ①格… ②沃… ③杨… ④边…
III. ①学业评定 IV. ①G424. 75

中国版本图书馆 CIP 数据核字 (2011) 第 152765 号

北京市版权局著作权合同登记 图字: 01 - 2009 - 2827 号

教育测量与评价丛书
学业成就评测 (第9版)
XUEYE CHENGJIU PINGCE (DI 9 BAN)

出版发行	教育科学出版社	市场部电话	010 - 64989009
社 址	北京·朝阳区安慧北里安园甲9号	编辑部电话	010 - 64989421
邮 编	100101	网 址	http://www.esph.com.cn
传 真	010 - 64891796		
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	2011 年 12 月第 1 版
印 张	16.5	印 次	2011 年 12 月第 1 次印刷
字 数	266 千	定 价	36.00 元

如有印装质量问题, 请到所购图书销售部门联系调换。

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，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是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，历经比纳（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，从经典测量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立和民办考试与评价机构，迄今已有近百年的历史。

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，以为借鉴。

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，每年有数以千万的人群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性的智力支持。

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中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*）和由亚普·希润斯（Jaap Scheerens）教授等人撰写的《教育评价与监测：系统的方式》（*Educational Evaluation, Assessment and Monitoring: A Systematic 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、技术的规范与要求。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。

前言

《学业成就评测》重点讨论了成就评测的原理与实践，以便实现更加有效的课堂教学和提高学生的学习效果。

有效教学包括：（1）用表现术语（performance term）清晰地表述对学生学习的期望；（2）设计与期望的学习成果相匹配的教学活动；以及（3）运用与预期学习成果和教学活动相协调的评测程序。本书集中于成就评测，但它把评测视做教学过程的内在组成部分。本书可帮助评价学习准备程度、监控并改进学习以及对学生成就进行评分或结业审查。如果设计评测与备课同时进行，则两者可共同发挥作用，以帮助学生达到教学所预期的学习成果。

前面2章描述了评测的过程和它在教学中的角色。第3章和第4章讨论了在设计评测时对教育目标分类法的使用和计算评测工具效度和信度的方法。其余的几章介绍如何准备测验和表现评测、如何用档案来评测、如何对学生成就进行评分和报告，以及如何解释标准化成就测验的分数。

第9版包括一些变化：

1. 增加了新的一章（第10章），它讨论了在表现评测（performance assessment）中如何安排环境、如何帮助学生和观察者做好必要的准备。
2. 删去了第3章，将内容标准在评测中的角色放在第1章中进行讨论。
3. 有关评测的效度和信度被移到书中更前面的位置，并且增加了计算效度系数的步骤方面的内容。
4. 对图和表格进行了编辑，以更清晰地说明在设计评测方法和工具时，如何使用《布鲁姆教育目标分类法》（*Bloom's Taxonomy of Educational Objectives*）的修订版。
5. 更新了几章的“参考文献与补充阅读”部分。

本书的写作方式简单而直接，经常使用例子和图示来说明要点。它对测验和

表现评测作了较为全面的讨论，以便两者能在最恰当的地方使用。总之，对于想学习如何设计有效的评测工具以促进有效教学和学生学习的的人而言，本书能够提供可操作性的帮助。

我们要感谢对本书提出宝贵意见的评审者：谢莉尔·康明斯（Cheryl Cummins），德尔塔州立大学；约翰·K. 梅斯（John K. Meis）博士，弗拉格勒学院；帕特里夏·I. 马尔卡希－欧内特（Patricial I. Mulcahy-Ernt）博士，桥港大学。

我们也要感谢本书中所提到的作者和出版社——艾琳和培根出版社（Allyn and Bacon）的编辑和其他所有对本书出版给予帮助的人。最后，要感谢我们的妻子——玛莉（Marie）和比莉·琼（Billie Jo）——在此次修改工作中给予我们的支持。

诺曼·E. 格伦隆德

C. 基思·沃

目 录

Contents

前言

第 1 章

成就评测与教学 / 1

教学与评测的关系 / 3

教学过程中的评测 / 4

评测促进学习的其他方式 / 10

学生评测中的教师标准 / 11

要点总结 / 13

参考文献与补充阅读 / 13

第 2 章

学生评测的类型 / 15

评测方法的主要类型 / 16

对学生进行有效评测的指导方针 / 19

常模参照和标准参照评测 / 23

要点总结 / 26

参考文献与补充阅读 / 27

第 3 章

评测的计划 / 28

使用分类法作为指导 / 29

	教学目标的角色 / 32
	评测的准备 / 36
	评测程序的种类 / 37
	评测计划中的效度与信度 / 38
	要点总结 / 41
	参考文献与补充阅读 / 42
第 4 章	效度与信度 / 44
	效度 / 45
	信度 / 57
	要点总结 / 65
	参考文献与补充阅读 / 66
第 5 章	设计和使用成就测验 / 68
	确定教学目标 / 69
	准备测验的细目表 / 71
	设计恰当的测验题目时的考虑 / 74
	题目设计的一般性指南 / 80
	在测验中题目位置的编排 / 82
	准备指导语 / 82
	测验的实施和评分 / 84
	分析测验题目的有效性 / 84
	要点总结 / 88
	参考文献与补充阅读 / 89
第 6 章	选择型题目的编制：选择题 / 90
	选择题的特点 / 90
	编制选择题的规则 / 92
	要点总结 / 108
	参考文献与补充阅读 / 109

第7章	编制选择型题目：是非题、匹配题和阅读理解题 / 110
	是非题 / 111
	匹配题 / 118
	阅读理解题 / 121
	要点总结 / 126
	参考文献与补充阅读 / 127
第8章	问答题的编制：简答题和论述题 / 128
	简答题 / 129
	论述题 / 133
	要点总结 / 143
	参考文献与补充阅读 / 144
第9章	表现评测 / 145
	操作技能的组成成分 / 146
	表现评测的目标陈述 / 147
	限制型和扩展型表现任务 / 148
	表现评测的准备步骤 / 149
	提高表现评测的现实度 / 171
	要点总结 / 172
	参考文献与补充阅读 / 173
第10章	表现评测的准备 / 174
	环境的准备 / 174
	学生的准备 / 175
	观察者的准备 / 177
	表现评测中的诊断性分析 / 178
	要点总结 / 179
	参考文献与补充阅读 / 179

第 1 章

成就评测与教学

学完本章后,你将能够:^①:

1. 解释为什么测验和表现评测在成就评测 (achievement assessment) 中都很重要。
2. 写出成就评测的定义。
3. 描述教学与评测之间的关系。
4. 辨别评测在教学过程中发挥的多种作用。
5. 列出评测对直接促进学生学习的多种作用。

在多年的学校教育中,我们都参加过各种各样的纸笔测验 (paper-and-pencil tests)。其中,一些是由教师设计的,要求我们选择一个答案 (例如,是非题、选择题或匹配题) 或给出一个答案 (例如,简答题或论述题)。另一些则是学业能力或成绩的标准化测验,主要采用选择题。纸笔测验在学校中应用广泛,这至少可以归于它们能高效地考查许多学习成果,而且评分和记录都很方便。

^① 由于排版空间有限,我们未能使用我们更喜欢的表述预期学习成果的方法,即本书第 3 章所描述的两步法 (two-step method)。尽管如此,上面的表述为你学习和应用各章内容提供了一个关注点。

- 2 近年来，这种过于强调纸笔测验的做法遭到了一些批评。有些批评者主张，应更重视评测真实的、“现实生活的”任务（例如解决现实世界中存在的问题）。另外一些批评者主张，至少应该用**备选类型评测**（例如口头表达、演示、档案评测等）来部分取代纸笔测验。有些批评虽然过于极端，但它们仍然强调了更多关注学生实际表现的重要性（参见框 1.1）。如果你想知道学生是否会写作，就让他们写点儿什么。如果你想知道学生是否会操作机器，就让他们操作一下那个机器。如果你想知道学生是否会做实验，就让他们做个实验。简而言之，如果你想知道他们是否能完成某项任务，就让他们去完成那个任务。毫无疑问，如果在学校里更重视表现评测，可以使我们更好地评测预期的学习成果（intended learning outcomes）。但是，即便我们强调要采用更直接的基于表现的任务（performance-based tasks），纸笔测验仍有其重要作用。

框 1.1 经常使用的评测术语

表现评测（Performance Assessment）

- 这种评测要求学生实际去完成一个或一系列任务（例如写故事、作演讲、做实验、操作机器等），以此来展现他们的理解程度以及所获得的技能情况。

备选类型评测（Alternative Assessment）

- 表现评测的一种说法，旨在强调是传统纸笔测验的一种替代形式。

真实性评测（Authentic Assessment）

- 表现评测的一种说法，旨在强调在理解基础上的应用和解决“真实世界”情境中的实际问题的技能。

- 3 大多数类型的表现都有一个对它而言很重要的知识要素。好的写作包括许多因素，比如在词汇、语法和拼写等方面的知识。用一个写作任务并不能很好地考查这些方面，因为我们往往只使用那些我们知道的词语，只使用我们能自如断句的句子结构，用我们会拼写的词语来代替我们不会拼写的。这样，在写作中，我们可以通过设计来隐藏我们的弱点。一个专门针对词汇、语法和拼写的测验可以发现这些弱点，进而帮助提高写作技能。只是不要把测验结果看做“写作能力”的体现。这些测验考查对写作有用的知识，而要判断写作能力，就要看实际的作

品（表现评测）。与此类似，就操作机器而言，实际操作机器是最终目标，但在实际动手进行表现评测之前，或许需要通过测验来考查学生在机器操作和所需遵守的安全规范等方面的知识。同样，在做实验之前，可以通过测验来判断学生在多大程度上了解一个控制良好的实验所需的各种参数。

贯穿本书的重点是成就评测，这既包括纸笔测验，也包括表现评测。测验可以为从简单到复杂的许多重要的学习成果提供一个直观的测量，它们也可以为评测和改进实际技能提供所需信息。因此，虽然我们应尽可能地使评测接近现实生活，在学校环境的局限下，我们既需要测验也需要基于表现的任务，以便完整地评测学生成就。

本书中，成就评测是一个宽泛的概念，包括用以判断学生在多大程度上达到了教学的预期学习成果的所有不同方法。因为我们只讨论成就评测，简便起见，本书只用评测这个词。

教学与评测的关系

在准备任何类型的教学项目时，我们的主要关注都是“怎样才能最有效地促进学生学习？”当我们思考这个问题时，我们很自然地将注意力转向教学方法和教学材料。但同时，我们也应当考虑评测在教学过程中的作用。如果设计合理、使用得当，评测手段可帮助实现更有效的教学和更好的学生学习。

教学和评测之间的密切关系可见表 1.1。两者都要求我们清晰明确地提出学生要达到的学习成果；设计出与有效教学相对应的良好评测所需具有的条件。这个关系凸显出拓宽备课（instructional planning）的范围，把评测计划（assessment planning）包含于内的重要性。那种典型的把备课仅局限于教—学过程的做法是不够的。要有效教学，我们就要把我们的关注拓宽到教—学—评测整个过程，把评测作为教学项目的一个基本组成部分。和所有的教学活动一样，评测的主要功能是改进学习，它可从多方面发挥此作用。

表 1.1 教学与评测的关系

教 学	评 测
具备下列条件时，教学最有效： 1. 直接指向一系列界定清晰的预期学习成果。 2. 教学方法和教学材料与所欲达到的学习成果相一致。 3. 教学设计与学生的特征和需要相匹配。 4. 教学决策的依据是有意义的、可靠的、相关的信息。 5. 定期告诉学生他们的学习进展情况。 6. 为那些没能获得预期学习成果的学生提供补救。 7. 定期审查教学有效性，根据需要修正预期学习成果和教学。	具备下列条件时，评测最有效： 1. 评测的设计针对一系列界定清晰的预期学习成果。 2. 评测的性质和功能与所要评测的学习成果相一致。 3. 评测的设计与相关的学生特征相匹配，并且对每个人都公平。 4. 评测提供有意义的、可靠的、相关的信息。 5. 及时给学生提供关于评测结果的反馈。 6. 利用评测结果，揭示出具体的学习弱点。 7. 评测结果为评价教学目标、教学方法和教学材料提供适当有用信息。

教学过程中的评测

要使评测与教学完全整合，评测计划应在备课过程中制订。从教学开始到结束，教师需要作出无数决策。通过提供更客观的信息作为判断的基础，仔细规划的评测步骤可以改进许多决策的有效性。让我们来考虑教师需要作出的一些决策：（1）在备课时；（2）在教学开始时；（3）在教学过程中；（4）在教学结束时。

备课

在开始建造一个房屋前，大量的思考和工作花在制作蓝图上。蓝图提供指导，也确定了房屋建成后应是什么样子。它像一个向导，在建造过程中保证质量。在设计一个课程（课堂材料、学生活动、课程计划等）之前，教师也需要

一个蓝图，以确定学生学完课程后看起来应是什么样子。教师需要回答下面两个问题：

1. 教学的预期学习成果是什么？
2. 我们如何知道学生是否获得了预期的学习成果？

在公共教育系统中，这些问题的答案是由学校所在的州通过制定**内容标准**（content standards）来提供的。内容标准描述了在某段学习（例如，一个年级或几个年级）结束时学生应当知道的东西或会做的事。它们为课程开发、教学和学生成就评测提供了一个框架。许多职业组织在它们各自的专业领域也制定了一系列的内容标准。人们希望，通过使用这些标准可以提高对学生的成就期望、改进公共教育的质量、造就更有知情力的公民，进而使我国比其他国家更有竞争力。

内容标准包括一些陈述，以大体确定学生应当学到什么。每个标准伴有几个**基线**（benchmarks），以澄清当学生达到内容标准时应知道的东西或应能做的事。下面的标准和基线给出了小学早期阶段的一个可能标准。

理解测量方法

1. 知道日常测量中使用的工具和单位。
2. 知道每个测量工具各种单位之间的关系。
3. 应用测量工具和单位解决测量问题。

如上例所示，许多标准用诸如**知道**、**理解**、**应用**等词语。因此，内容标准通常不是在教学目标（instructional objective）的层次上制定，所以也不会描述学生如何证实他们获得了学习成果。在制订相应的教学目标时，我们需要用明确的词语来澄清**知道**（例如，识别、描述）、**理解**（例如，解释、举例）、**应用**（例如，计算、转换）等词的含义。下列的教学目标或许可以把上面示例的标准变成教学和评测可用的形式。

1. 理解测量方法

- 1.1 识别测量距离、时间、重量、温度和体积时所用的工具和单位。
- 1.2 描述每个测量工具各种单位之间的关系（例如，英寸、英尺和码）。
- 1.3 对于某种给定的测量，从一种单位转换到另一种（例如，从英寸到英

尺)。

1.4 在给定情境下测量某个值 (例如, 线段长度)。

1.5 说出日常生活中使用每种测量工具的例子。

1.6 用适当的工具和单位解决测量问题。

- 6 标准被转化成教学目标后, 它们为规划教学活动和制作评测工具提供了指导 (蓝图)。例如, 有了上面的目标 1.1, 教师将被引导着去设计一些活动和评测项目, 以要求学生识别在测量长度、时间、重量、温度和体积时所用的正确的工具和单位。关于如何写出用于教学和评测的教学目标, 参见格伦隆德和布鲁克哈特 (Gronlund & Brookhart, 2009) 的文章。

教学开始 (安置性评测)

在进行教学前, 教师需要回答下面两个问题:

1. 对开始教学所必备的技能 and 能力而言, 学生已经掌握了多少? 在多大程度上具有开始教学所需的技能和能力?
2. 对教学所预期的学习成果而言, 学生已经掌握了多少? 在多大程度上已经达到了教学计划的预期学习成果?

关于第一个问题的信息通常通过准备性前测 (readiness pretests) 获得。这种测验在课程或教学单元开始时进行, 考查那些为达到教学目标所必须预先具备的技能。例如, 在几何课开始时, 可能会进行计算技能测验; 在德语课开始时, 可能会进行英语语法测验。对于缺乏预备技能的学生, 可以给他们补课, 或者把他们安排在需要较低预备技能的特别组。

回答第二个问题, 通常采用安置性前测 (placement pretests)。这种测验考查教学计划的预期学习成果。虽然它很可能和教学结束时的试卷相同, 更好的情况下, 它应换一种形式。通过测试, 我们可以了解, 学生是否已经掌握了 we 打算在教学中包括的一些材料。如果是, 我们或许需要修改教学计划, 鼓励一些学生跳过某些单元, 并把一些学生分到更高一级的教学。图 1.1 概括了安置评测的功能。

除了使用前测外, 基于表现的任务对于判断学生的初始技能 (entry skills) 或许也有帮助。例如, 在写作课上, 在教学开始时收取写作样例 (writing sam-