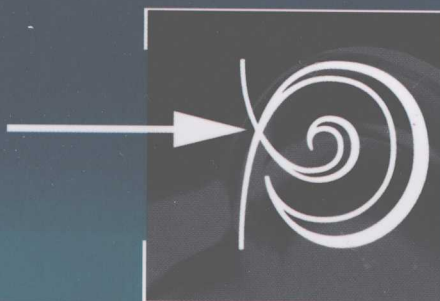




Assessment of Student Achievement



教育科学精品教材译丛



学业成就测评

(第七版)

[美] 诺尔曼·E. 格兰伦德 著

罗黎辉 等译
孙亚玲

凤凰出版传媒集团
 江苏教育出版社
JIANGSU EDUCATION PUBLISHING HOUSE





教育科学精品教材译丛

Assessment of Student Achievement

学业成就测评 (第七版)

[美] 诺尔曼·E.格朗伦德 著

罗黎辉 等译
孙亚玲

凤凰出版传媒集团



江苏教育出版社

JIANGSU EDUCATION PUBLISHING HOUSE

著作权合同登记图字: 10-2004-163 号

图书在版编目(CIP)数据

学业成就测评(第7版)/(美)诺尔曼·E. 格朗伦德著/罗黎辉, 孙亚玲译. —南京: 江苏教育出版社, 2008. 4
ISBN 978-7-5343-8352-6

I. 学… II. ①诺…②罗…③孙… III. 高等教育—教育理论—研究 IV. G640

中国版本图书馆 CIP 数据核字(2008)第 056155 号

Norman E. Gronlund

Assessment of Student Achievement (Seventh Edition)

ISBN: 0-205-36610-4

Copyright © 2003, 1998, 1993, 1988, 1982, 1977, 1968 by The Pearson Education, Inc.
Original language published by The Pearson Education, Inc. All rights reserved. No part of this publication may be reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.
Simplified Chinese translation edition jointly published by Pearson Education and Jiangsu Education Publishing House.

本书中文简体字翻译版由江苏教育出版社和培生教育出版集团合作出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有 Pearson Education 公司防伪标签, 无标签者不得销售。

江苏省版权局著作权合同登记号: 10-2004-163

教育科学精品教材译丛

学业成就测评(第七版)

Assessment of Student Achievement (Seventh Edition)

[美] 诺尔曼·E. 格朗伦德 著

罗黎辉 孙亚玲 等译

责任编辑 韩宇新

出版发行 凤凰出版传媒集团

江苏教育出版社

(南京市马家街 31 号, 邮政编码: 210009 网址: www.1088.com.cn)

经销 江苏省新华发行集团有限公司

照排 南京展望文化发展有限公司

印刷 江苏淮阴新华印刷厂(厂址: 淮安市淮海北路 44 号 邮编: 223001)

开本 787×1092 毫米 1/16 印张 12.25 字数 250 000

2008 年 4 月第 1 版 2008 年 4 月第 1 次印刷

印数 1-5 000 册

ISBN 978-7-5343-8352-6

定价 24.00 元

批发电话 025-83260760, 83260768

邮购电话 025-85400774, 8008289797

短信咨询 10602585420909

E-mail jsep@vip.163.com

盗版举报 025-83204538

苏教版图书若有印装错误, 可向承印厂调换。

提供盗版线索者给予重奖。

编 委 会

顾 问

顾明远 章新胜

主 编

朱永新

副主编

严文蕃 张胜勇

编委(按姓氏笔画为序)

王智新 卢乃桂 许庆豫 朱小蔓 吴康宁

张斌贤 周 川 俞慧洵 赵 明 赵中建

钟启泉 徐 辉 袁振国 董 奇

James Campell Thomas Shuell

海外咨询委员会

主任委员

韦 钰

委员(按姓氏笔画为序)

万毅平博士…美国肯尼索大学教育学院院长、教授

马立平博士…美国卡内基教育基金会

关小茹博士…美国芝加哥德保罗大学教学科技部主任

孙 静博士…澳大利亚昆士兰科技大学早期儿童应用研究中心

杨效斯博士…美国芝加哥森林湖学院亚洲研究中心主任

陈欣银博士…加拿大西安大略大学发展心理研究室主任

周 正博士…美国纽约圣约翰大学心理学系

秦志宁博士…美国明尼苏达州哈普金斯教育局测量评估部主任

彭凯平博士…美国加利福尼亚州立大学教授

蓝 云博士…美国得克萨斯州工科大学教育学院副院长

总序

作

为高校教师,我们中的许多人常常为教育科学教材的陈旧落后而痛心疾首;作为教育学人,我们中的许多人也常常对经济学、社会学等显学学科教材建设的突飞猛进而称美不已。

于是,我们坐卧不安,我们摩拳擦掌,我们立志超越,我们走到了一起。经过几年的努力,涵盖当代高等学校教育学专业的全部主干课程的大型海外教材《教育科学精品教材译丛》(下面简称《译丛》)终于呈现在读者面前。

许多年来,我国高等师范教育和高等学校教育学专业课程改革的步伐极为缓慢,师范教育的教育学、心理学、教材教法这三门课程多年不变,教育学专业的课程内容陈旧,课程的选择空间相当狭小。可以说,改变高等师范教育课程和高等学校教育学课程的落后状况,是《译丛》最为基本的宗旨。

另一方面,随着教育事业改革的深化,教育实践中产生的问题日益复杂,解决这些问题需要极为丰富的教育科学知识和能力。《译丛》追求的另一宗旨正是通过奉献世界上最先进的教育科学知识体系,促进我国教育事业改革的深化。

在过去的几年中,高等学校课程改革已经取得了相当明显的成效。深化课程改革的一条重要途径是引进国外尤其是发达国家的高校教材,藉此提高教育质量和增进学生的学习能力。《译丛》的宗旨和思路与我国高校教材改革的这种方向是一致的,而且是高校教材改革过程的组成部分。

促进学术交流,是《译丛》向往的又一宗旨。学术沟通的障碍,表征是交际语言,而深层原因则是学术语言与学术规范。《译丛》希望通过引进国外的教育科学知识体系和贯穿其

中的研究方法与表达方式,促进我国教育科学学术事业的进步,并为其走向世界奠定基础 and 开辟道路。

《译丛》是建国以来从海外引进的规模最大、门类最全的教育学科教材,被国内媒体称为“又一次重要的拿来主义”。在科教兴国的基本国策背景下,它所蕴涵的巨大社会意义已经超出教材本身。因此,《译丛》的编委会和出版者——江苏教育出版社对此高度重视,并为此做了大量的细致而扎实的工作。第一,组建了强大的编委会和翻译队伍。《译丛》的编委会阵容整齐,有各师范大学的博士生导师、教授以及一批海外教育专家;主要翻译人员和审校者均是教育科学专业的博士或教育科学领域的教授,其中一些译者长期旅居国外,并从事教育科学专业的研究和教学工作,他们均在教育科学领域具有相当深厚的积累,可以确保《译丛》的翻译质量。第二,精心筛选选题。《译丛》的入选图书品质上乘,所有选题皆经中、日、美等国专家反复磋商论证,精选而成。其中一些书目为国外学术机构所推荐,在国外大学拥有广泛的学术声誉。许多教材一版再版,最多的已达第十五版。

我们希望,这套教材能成为国内教育科学的替代课本或重要参考书,同时也能作为各地教师继续教育的重要资料。

我们期待,这套教材能给中国教育理论界带来一些观念和方法上的启示,为我国的教育科学的教学和研究,尤其是教材编写工作提供一定的借鉴。

我们相信,这套教材会得到许多中小学教师、校长、教育行政机关干部、教育科学研究人员、教育专业的研究生以及高校在校学生的关注和选用。

当然,我们更希望、更期待的是创新和超越。希望和期待我国的教育科学工作者编写出高水平的、具有中国特色的教材。站得更高才能看得更远,看得更远才能做得更好,希望我们这套教材能使中国教育理论界有一个更高的起点,使中国的教师和师范学生有一个开阔的眼界。需要说明的是,原书附有大量的各类索引,但为降低图书成本,减轻读者负担,我们只好割爱,敬请诸君谅解。

我们欢迎各种形式的参与和合作,欢迎专家和读者随时为我们荐书,随时提出各种建议和评论。

《教育科学精品教材译丛》编委会

二〇〇二年四月

前言

《学

业成就测评(第七版)》是一本重点介绍如何对教学的所有预期学习结果(简单或复杂)进行测评的实用指南。书中将测评定义为笔试和实作测评的总称,两种方法各用于最合适的场合。全书的前提是:高效的教学必须进行有效的测评,而且教学和测评的最终目标都是提高学习效果。

全书以课堂测验与实作测评的准备与使用的章节开头,接下来的一章是关于评分与反馈。最后两章介绍了如何理解标准化测试分数,以及效度与信度的问题,展现了测评中更具技巧性的方面,不过这些概念和过程都很容易掌握,不需要任何该领域的预备知识。它们被安排在最后,这样就可以在最适合课程进度的地方加以使用。第二章简要描述了效度与信度的概念,还列举了测评工具准备阶段构建效度与信度的各种方法,便于想在课程中使用技巧手段的人使用。

最新版有很多改变:

1. 增加了关于延展性实作测评的章节(第八章),侧重介绍对综合复杂的实作结果的测评。
2. 增加了关于档案测评的章节(第九章),描述如何准备、使用并评估课堂档案。
3. 第三章扩充为制定测评计划,扩充了有关学业成就测试准备的内容,囊括了测试准备的方方面面。取消了关于召集测试、实施测试和评估测试的章节,因为第三章已经涵盖了绝大部分相关内容。
4. 修改了以前版本中有关实作测评的章节,新版中该部分仅侧重描述对技术与作品进行实作测评的传统方法。

5. 扩充了有关评分的章节,新增了向学生和家长反馈分数的内容。
6. 各章节新增了诸多成套式表格,列出进行准备、使用实作测评及评分的各个步骤。
7. 各章的补充阅读中增加了新的参考文献。

现在,本书将重心平均分配在笔试及实作测评两方面,并说明其在教学中的重要作用。介绍相关材料的方法简单、直接、易懂,但并没有因此而轻视基本概念或牺牲技术准确性。书中提供了很多实用案例。此外,检核表、成套式表格及各章的要点总结也有助于学习相关内容。学习者不需要具备测量或统计学方面的知识就能理解本书内容。简而言之,这是初学者的实用指南。

我要感谢西南奥克拉荷马州立大学的校订员詹姆斯·泰特(James Tate),还有弗吉尼亚理工大学的简·亚伯拉罕(Jane Abraham),感谢他们付出的时间和反馈意见。

我还要向书中所参阅的作者及出版社致谢,感谢艾琳和培根(Allyn and Bacon)出版社的所有编辑人员,感谢艾琳·帕娜默(Irene Plamer)出色的打印工作,感谢我的妻子玛丽(Marie)在本书修订过程中的耐心与理解。

诺尔曼·E. 格朗伦德(N. E. G)

目 录

总序 / i	
前言 / i	
第一章 成就测评与教学 / 1	
教学与测评的关系 / 2	
教学过程中的测评 / 3	
教学开始阶段(配置性测评) / 3	
教学过程中(形成性与诊断性测评) / 5	
教学结束阶段(终结性测评) / 6	
测评促进学习的其他方法 / 7	
学习动机 / 7	
学习效果的保持与迁移 / 7	
学生的自我测评 / 8	
教学效果评价 / 8	
学生测评对教师的要求 / 8	
要点小结 / 9	
第二章 学生测评的实质 / 11	
测评方法的主要类型 / 12	
测评任务的真实性 / 12	
测试任务的复杂性 / 13	
测评所需的时间 / 13	
评分的主观性 / 14	
有效测评学生的指导原则 / 14	
测评计划的效度与信度 / 18	
常模参照测评和标准参照测评 / 20	
要点小结 / 22	
第三章 测评方案的制定 / 24	
预期学习结果的类型 / 25	
教学目标的作用 / 26	
表述教学目标 / 26	
确定样本目标的辅助资料 / 27	
准备与使用学业成就测评 / 28	

详细说明教学目标 / 29	
用具体的语言界定一般性的学习结果 / 29	
制定考试内容规范 / 31	
制作考试内容规范表 / 31	
编写试题时需要考虑的因素 / 33	
选择将要使用的考试题目类型 / 33	
使题目与具体的学习结果相一致 / 35	
提高题目的功能性内容 / 36	
确定试题的数量 / 37	
编写题目的一般性指导原则 / 38	
组题 / 39	
编写考试指令 / 40	
猜答案的问题 / 40	
复查和评价组好的试卷 / 41	
实施考试及评分 / 42	
分析试题的有效性 / 42	
简单的题目分析程序 / 42	
要点小结 / 45	
第四章 编制选择题：多项选择题 / 46	
多项选择题的性质 / 46	
多项选择题的使用 / 47	
知识类题型 / 48	
理解类题型 / 49	
应用类题型 / 51	
选择题的命题原则 / 52	
要点小结 / 65	
第五章 编制选择题：是非题、配对题与解析题 / 66	
是非题 / 66	
是非题编写原则 / 68	
配对题 / 71	
配对题命题原则 / 73	
解析题 / 74	
解析题命题原则 / 76	
要点小结 / 78	
第六章 编制主观题：简答题和论述题 / 79	
简答题 / 79	
简答题命题规则 / 81	

论述题 / 82

论述题的类型 / 83

论述题命题规则 / 85

论述题评分规则 / 87

要点小结 / 90

第七章 传统的技能与技能发展水平实作测评 / 91

实作技能的要素 / 92

制定实作测评的目标 / 92

限制性的实作任务和延展性的实作任务 / 93

准备实作测评的步骤 / 94

说明实作的结果 / 94

选择测评的重点 / 97

测评实作过程 / 97

测评实作成果 / 97

选择恰当的真实度 / 98

选择完成实作的场景 / 99

笔试实作 / 99

鉴别测试 / 100

有条理性的实作测试 / 100

模拟实作 / 102

实作任务样本 / 103

延展性研究项目 / 103

选择观察、记录和评分的方法 / 103

系统观察和逸事记录 / 103

检核表 / 104

评定量表 / 105

分项评分与整体性评分 / 106

完善实作测评 / 108

要点小结 / 108

第八章 延展性实作测评 / 110

扩展传统实作测评的方法 / 111

进一步强调认知技能的作用 / 111

提高交际技能的作用 / 112

增加视觉材料的使用 / 112

扩大学生的参与面 / 112

提高任务的真实性 / 112

增加任务的复杂性 / 113

将实作任务本身作为实作的提示 / 113

设计实作任务的指导性原则 / 114

制定测评实作任务的质量标准 / 115

利用学生的研究项目进行实作测评 / 117

建立质量标准和水准 / 117

选择与陈述问题 / 118

查找和选取资料 / 119

撰写研究报告 / 119

设计并完成研究项目或研制出成果 / 121

对研究项目作口头陈述和答辩 / 121

评价学生的研究项目 / 122

准备评定量表 / 122

准备整体评分指标 / 122

要点小结 / 123

第九章 档案测评 / 125

使用课堂档案的优点 / 126

使用档案测评的计划 / 127

档案的目的 / 127

归档的条目种类 / 128

选择和评价档案条目的原则 / 128

档案的维护与使用 / 129

对档案的评价 / 130

对档案构建的评价 / 130

评价学生总体表现的档案 / 131

评价学生的进步 / 131

评价学生最终的实作水平 / 132

要点小结 / 134

第十章 等级评定与报告 / 135

选择评定等级的依据 / 135

绝对评级 / 136

相对评级 / 136

学习能力、学习进步和学生的努力 / 138

绝对评级和相对评级的综合运用 / 139

综合数据来评定等级 / 140

有效、公正地评定等级的原则 / 142

向学生和家长报告 / 143

使用档案 / 144

要点小结 / 145

第十一章 标准化成绩测验分数的解释 / 146

标准化成绩测验的特征 / 146

常模参照分数的解释 / 147

百分等级 / 148

百分区间 / 150

年级等价成绩 / 151

标准分 / 152

平均分、标准差和正态曲线 / 153

z 分数 / 154

T 分数 / 155

正态曲线等价分数(NCE) / 156

能力分数 / 156

九段分数 / 156

标准参照解释 / 157

要点小结 / 158

第十二章 效度和信度 / 160

效度 / 160

内容关联指标 / 162

校标关联指标 / 163

构想关联指标 / 165

使用评价结果的意义 / 167

信度 / 167

评估测验分数的信度 / 168

评估成绩评价的信度 / 173

要点小结 / 174

词汇表 / 176

参考文献及补充读物 / 179

第一章 成就测评 与教学

本章使读者能：^{〔1〕}

1. 理解为什么考试和实际操作测评在成就测评中均很重要。
2. 对成就测评作出界定。
3. 阐述教学与测评的关系。
4. 认识教学过程中测评的各种作用。
5. 系统把握测评直接促进学习的各种途径。

凡受过一些年教育的人，都曾经参加过各种笔试，有的是教师自编测验——其试题或要求选择作答（如：是非题、多项选择题或配对题）或要求具体作答（如：简答题或论述题）；有的是标准化能力倾向测验或标准化成绩测验——其试题主要采用多项选择题。笔试之所以在学校通行，至少，部分原因是基于对效率的考虑。大部分学习结果都可以用笔试方式进行测试，而且，笔试易于评分，评分结果也易于记载和保存。

近年来，出现了反对过多注重笔试的倾向。一些对笔试持批评态度的人，主张更重视真实性测评（authentic assessment），用“现实生活”作业（譬如，解决现实社会的实际问题）进行测评。另一些批评者则认为，笔试应该被淘汰，至少应部分地为替代性测评（alternative assessment）所取代。有些批评虽过于极端，但是，它们突显了重视学生实践能力的必要性（详见专栏 1-1）。若要判断学生的写作能力，得让学生写作；若要判断学生的机械操作能力，得让学生操作机器；如果要判断学生的实验能力，就得让他们做实验。总之，要想知道学生某方面的实作能力，就得让他们去进行相关的实作。无疑，把实作测评（performance assessment）摆上学校教育更重要的位置，有利于促进预期学习结果测评的发展。但是，在教学中，无论我们怎样重视实作测评，笔试依然具有重要的作用。

〔1〕 虽限于篇幅，不能采用在第三章推崇的两步法来阐述预期的学习效果。但是，下述内容可为大家的学习，以及为各章内容的应用提供一个抓手。

专栏 1-1 常用评价词汇

实作测评	要求学生通过实际完成一项或一组作业(如,撰写故事、发表演说、开展实验或操作机器),以证明他们所取得的认识成就和技能成就的测评。
替代性测评	实作测评的一种,强调其测评方法有别于传统笔试。
真实性测评	实作测评的一种,重视用知识和技能解决现实中的实际问题,强调重视这一点的重要意义。

大多数实践教学活动均需要知识的支撑,知识对实践至关重要。要写出好文章,离不开词汇、语法和拼写等方面的知识。一篇文章,并不能很好地反映写作知识水平的高低。因为,在写作中,我们一般只使用自己知道的词汇和得心应手的句型。在碰到写不出来的词汇时,我们往往会用其他能写得出的词汇来替代。所以,在写作中,我们可以通过精心构思来掩盖自己语言上的不足。但是,独立的词汇测验、语法测验或拼写测验却能够发现这方面的不足,以利于学生提高写作能力。不过,我们还不能把这些测验解释为“作文能力”测验。它们测量的是写作中有用的知识,写作能力要通过实际写作来测评(实作测评)。操作机器同样如此,学会实际操控机器是最终的目的。但在着手进行实作测评前,有必要先测试学生操作机器的知识及操作机器应遵守的安全防范措施。此外,在实验前,可以用测试手段了解学生是否具备做好一项实验必备的知识。

本书所要探讨的核心内容是学业成就测评,既包括笔试,又包括实作测评。考试可以直接对许多学习结果进行测量,从简单到复杂的学习结果都在其内。并且,考试还可以为实作测评及其改进提高实作质量,提供所需的信息。所以,尽管我们要尽可能为实现真实性测评而努力,但限于学校的环境条件,要全面地测评学业的成绩,考试和实作测评依然不可或缺。

本书涉及的**学业成就测评**(achievement assessment)是一个宽泛的范畴,它包括用于确定教学中学生在多大程度上取得预期学习效果的各种方法。由于我们现在重点关注的是学业成就测评,出于方便的考虑,全书仅用“测评”这一术语,不使用其他同义或近似术语。

教学与测评的关系

制定教学方案时,我们最关心的是“怎样才能最有效地促进学生的学习?”。面对这一问题,我们的注意力自然地集中到教学内容和教学方法之上。在这个时候,我们应该考虑教学过程中测评的作用问题。只要设计到位,运用得当,测评过程能够对提高教学质量、促进学生学习起到积极作用。

表 1-1 揭示了教学与测评之间的密切联系。我们不仅需要明确学生在教学中应

达到的教学目标;而且,需要通过精心的设计,准备好与高质量教学特性相匹配的测评。教学与测评的关系,使我们更清楚地认识到拓宽教学设计,纳入测评设计的重要意义。传统意义上,教学设计仅仅被局限于教—学的过程上。这是不够的。有效的教学,要求我们关注教—学—测评的过程,把测评作为教学计划的基本组成部分。教学测评的主要功能是,与其他教学活动一起,共同提高学生的学习效果。教学测评能够以多种方式为提高学生的学习效果发挥积极的作用。

表 1-1 教学与测评的关系

教 学	测 评
满足下列条件时,教学将最有效:	满足下列条件时,测评将最有效:
1. 以一系列界定清晰的教学目标为导向。	1. 以一系列界定清晰的教学目标为导向。
2. 教学内容与方法符合所要达到的教学结果的需要。	2. 测评的功能及实质与测评结果要相符。
3. 教学设计符合学生的特点和需要。	3. 测评的设计符合相关学生的特点,并且,对每个学生都公平合理。
4. 教学策略建立在有意义的、可靠的和相关的信息基础上。	4. 测评提供有意义的、可靠的和相关的信息。
5. 定期让学生了解自己的学习效果。	5. 及早向学生反馈测评结果。
6. 对尚未达到教学预期目标的学生采取补救措施。	6. 测评结果要能反映学生学习上的弱点。
7. 定期检查教学的效果,并对教学目标及教学活动进行必要的调整。	7. 测评结果要为评价教学目标、教学方法及教学内容提供有用的信息。

教学过程中的测评

为了与教学更好地相结合,测评的设计应与教学设计一并考虑。教师在教学的整个过程中,要作出一系列的教学决策。测评过程设计得好,能为教师作出基本判断提供较客观的信息,从而保障多数教学决策的有效性。我们来看看在教学开始阶段、教学过程中和教学结束阶段教师需要作出的一些决策。

教学开始阶段(配置性测评)

在新的教学开始前,教师需要回答两个主要问题:

1. 学生在多大程度上具备进行后续教学所需的知识与能力?
2. 学生在多大程度上已经达到了后续教学预期的教学目标?

第一个问题,通常以“预备状态”预测作出分析。预测一般安排在一门课程或一个单元的教学开始前进行。测验内容的范围应覆盖后续教学必需的各种技能。如在新开代数课时,应进行计算能力的测试;同样,在新开德语课之前,应进行英语语法的测试。对缺乏必备条件的学生应进行补习,或者把他们安排到预备状态较差的学生组别中去。

第二个问题,一般用“配置性”预测来寻求答案。配置性预测的内容应覆盖后续教学的各个预期目标。这与终结性教学测验正好相同。当然,配置性预测与终结性测验在测验形式上最好互不相同。在配置性测验中,我们要判断的是,学生是否已经掌握了我们纳入教学计划的内容。如果已经掌握,我们有必要调整教学计划,鼓励一部分学生跳过某些单元、把一部分学生配置到高起点的教学班中。配置性测评的功能如图1-1所示。

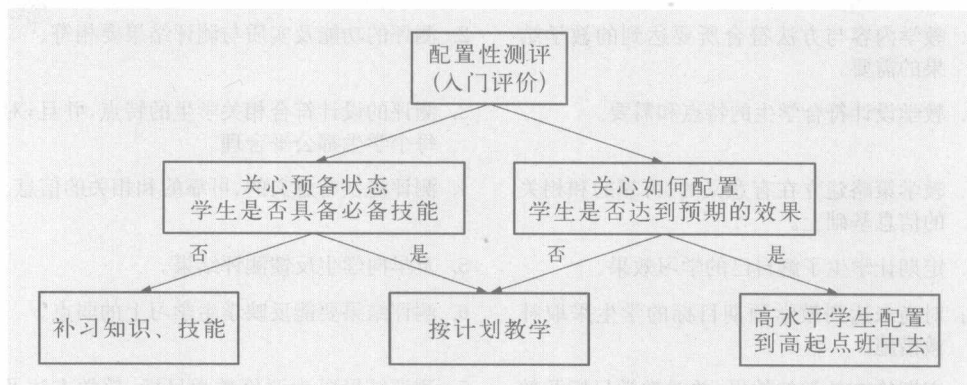


图 1-1 配置性测评功能简化模式图

除采用预测外,实作作业在判别入门技能上,也大有用武之地。譬如,在写作方面,我们可以在教学开始时抽取写作样本,建立一个基线,以此来测评学生今后写作水平的发展情况。如果在教学过程中我们把学生的作品作为档案保存下来,那么,据此作出的预测评,其价值将非比寻常。

预测评对教学促进作用的大小,取决于教学的性质,取决于对学生了解的程度,还取决于如何使用测评结果等。在新开代数课程时,算术预测可能颇为有效,但对于必备技能缺乏清晰界定的课程(如社会研究)来说,预测则可能毫无价值。同样,对不熟悉学生情况的新教师来说,基本技能测试对教学可能很有价值,但对深谙学生背景的老教师来说,意义则不大。另外,只有在打算查缺补漏,打算把学生安置到教学进程中最有利的位置,或者打算建立评价学生后续发展的基线时,预测评对教学才能起到促进作用。总之,要使预测评对教学产生最佳的效果,在进行教学设计时就要把它纳入