

英语课堂教学活动指南丛书 04

英语课堂 教学评价活动

凤凰出版传媒集团 译林出版社

外方主编 Paul Seligson

中方主编 林立

副主编 汪晓阳

Andy Baxter 著



英语课堂教学活动指南丛书 04

外方主编 / Paul Seligson

中方主编 / 林立

副主编 / 汪晓阳

[英] Andy Baxter / 著

司 洋 / 译

英语课堂教学评价活动

Evaluating Your Students

凤凰出版传媒集团
译林出版社

图书在版编目(CIP)数据

英语课堂教学评价活动/(英)巴克斯特(Baxter, A.)
著;司洋译. - 南京:译林出版社, 2008.1

(英语课堂教学活动指南丛书)

ISBN 978-7-5447-0237-9

I. 英… II. ①巴…②司… III. 英语课 - 课堂教学 - 教
学评议 - 中小学 IV. G633.412

中国版本图书馆 CIP 数据核字(2007)第 145682 号

Copyright © Andy Baxter 1997

© Richmond Publishingtm

Chinese Edition Copyright © Yilin Press 2007

All Rights Reserved.

本系列图书由英国Richmond Publishing出版公司授权译林出版社在中华人民共和国境内
(不包括中国香港、澳门特别行政区和中国台湾地区)独家出版发行。

登记号 图字: 10-2007-170号

“英语课堂教学活动指南”丛书由北京行走出版咨询有限公司策划。

版权所有, 翻印必究!

书 名 英语课堂教学评价活动
作 者 Andy Baxter
封面设计 侯海屏
排 版 刘红运
责任编辑 唐晓萌
特约编辑 裴凤敬
出版发行 凤凰出版传媒集团
译林出版社(南京湖南路 47 号 210009)
电子信箱 yilin@yilin.com
网 址 <http://www.yilin.com>
集团网址 凤凰出版传媒网 <http://www.ppm.cn>
印 刷 扬中市印刷有限公司
开 本 787 × 1092 1/16
印 张 7
版 次 2008 年 1 月第 1 版 2008 年 1 月第 1 次印刷
标准书号 ISBN 978-7-5447-0237-9
定 价 14.00 元

译林版图书若有印装错误可向承印厂调换

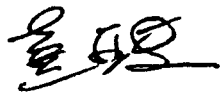
丛书序

外语课程的改革可以分为三个层面：一是宏观的、战略性的计划，如确定教学理念与设计课程目标、评价目标等；二是策略性层面（Candlin, 1984），即根据不同的地区差异与教学条件设计不同的教学实施大纲，其中包括对教学内容进行选择与排序，选择适当的材料，以及根据不同的教学需求采用适当的教学方法和教师培训等；最后是操作层面（Markee, 1997），即教师与学习者如何根据班级的学生实际情况进行日常教学的活动，包括材料的补充与选择。

可以说，自外语新课程实施以来，课程的教育教学理念、课程目标、教学材料等的确与以前的大纲有所不同，但是，教学方法、教学资源、教师职业发展等方方面面策略性层面与操作层面的变化却没有人们期待的那么明显。课程改革的理念和原则，经过几年的宣传和培训，虽然已经广为人知，但如何在教学实践中全面落实这些理念与原则，成为各级教育行政部门和外语教师面临的大问题。我个人认为，在课程实施中，最困难的还并不是教师教育教学理念的转变，而是各级教育部门、教师培训者，包括相关的教育出版部门在课程策略性层面、操作层面上多为一线教师想一想，他们在课程实施中最缺乏什么。一线教师是真正在比赛场地上奔跑的核心队员（key players）。他们需要什么样的营养和技能？他们需要什么样的技术支持？

当然，有关教育教学的理论培训与读本是需要。但是，教师们可能更需要的是比较方便、实用的教学资源；是稍加修改，甚至不用修改就可以用于教学活动的各种各样的辅助资料。教师们何尝不想采用更有效的教学方法？他们何尝不想使学生更喜欢学习？但是，巧妇难为无米之炊，如果没有方便的资料，有几个教师能凭空自己编写适当的、地道的外语辅助材料，设计语言得体、情境真实的活动内容？作为一名英语教材编写者，我深知教学活动材料设计的困难。这些困难不仅是理论水平上的，还有跨文化交际知识、学生心理状况、课堂教学过程的掌控，以及时间、精力、条件和创造力等诸多其他方面的因素。

人们也越来越清楚地认识到，教学的成功与否，绝不是仅仅靠什么这个教学法、那个教学法可以决定的。教学涉及的问题太复杂了，时间、学习环境、家庭环境、师资、接触语言的频度、心理定位、动机、态度、个人的不同学习风格等等，都可能影响教学效果。而其中一个重要的因素，就是教学材料的资源。“英语课堂教学活动指南”丛书和“英语语言教学实用技巧”丛书正是这样内容丰富、实用的外语课堂教学资源库。它们呈现了当今英语教学的许多核心话题，涵盖了教学分析、教学技能、教学方法等外语教师需要重点关注的方面。其贯穿始终的特点是实用，易于操作。相信这两套丛书定会受到教师们的欢迎，因为它们不仅有助于教学的发展，更会有助于教师的专业发展。教师的专业发展不仅有助于解决目前的资源短缺，而且还“授之以渔”，使教师个人的能力得以提升，从而使全国中小学英语教师水平上升到一个新的档次。



2007年7月

策划人语

外语课程改革自实施以来，我国基础外语教育领域发生了许多深刻的变化，影响到教学理念、教学方法、教学资源、教师职业发展等方方面面。这些变化都以前所未有的程度和速度改变着外语教学的现状，可以说是新中国成立后外语教育领域革命性的大变革。

面对这场大变革，相信每位外语教师都在以各种方式思考并参与。课程改革的理念和原则，经过几年的宣传和培训，已经深入人心。但如何在教学工作中全面落实这些理念与原则，成为外语教师面临的大问题。理论和实践相结合，从来都不是一件一蹴而就的事情，需要实践者反复思考、摸索。何况，在我国多年来形成的应试教育的情形毕竟对教师教学思想和实践的干扰与束缚根深蒂固。这就要求广大外语教育研究者和实践者努力探索出一条适合当前教学实际的创新之路。“英语课堂教学活动指南”丛书的出版就是这样的一种探索。

“英语课堂教学活动指南”丛书是一个内容丰富实用的外语课堂教学活动资源库。它的编写宗旨是给外语教师提供有效的教学资源，帮助其形成自己的教学模式，并以研究教学为乐事。

“英语课堂教学活动指南”丛书虽然是为中小学英语教师编写的，但是也适用于其他外语语种的教师、成人外语培训教师、希望从事外语教学工作的大学生和外语教师培训者。书中所有的课堂活动都是为水平较低的学生（从初学者到中低水平者）而设计的，因为这些学生在中小学中占多数。然而，大部分活动都可以轻而易举地改编成较高水平的学生可用的活动。

除了正文部分，“英语课堂教学活动指南”丛书的每一本还包括：

- 可复印活动页。这些内容可直接用于课堂教学（有些需要稍做修改），其中有的是供整个学年使用的，例如：评价记录表、项目计划表等。
- 教学技能培养任务。这些任务旨在使书中涉及的教学方法更具操作性，使教师对所读内容进行反思教学，帮助教师在课堂中尝试新的方法。
- 术语表。该丛书的主要原则之一是使用便捷。作者尽量避免使用行话或费解的专业术语。对于不可避免的术语，作者会在书后的术语表中进行解释。

真诚希望“英语课堂教学活动指南”丛书的出版能为广大外语教师的教学工作助一臂之力。

目 录

前 言	评价中的问题	1
第一部分	评估、测试和评价	4
第一章	为什么评估学生的学习	4
第二章	测试、教学和评价	7
第三章	评估什么	14
第四章	测试：好试卷的要素	18
第五章	测试和评价的方式	31
第二部分	为评估收集数据	35
第六章	语法和词汇的测试技巧	35
第七章	阅读和听力的测试技巧	44
第八章	开放式问题的测试技巧	50
第三部分	评估	51
第九章	评估口语和写作能力	51
第十章	学习过程评估和学习态度评估	60
第四部分	平时的评估	73
第十一章	连续性评估	73
第十二章	形成性评价	77
第十三章	总结性评估：面对面评估和行为评估报告	81
结束语	该改变评估方法了	92
附录一	可复印活动页	95
附录二	术语表	102
附录三	拓展性阅读书目	104
附录四	答案	104

前言

评价中的问题

“这本书为谁而写？”

作为教师，我们的任务之一就是评估学生。也就是说，我们需要评判学生学习的好坏。因此，任何一位教师都会从有关测试的书籍中获益。

如果你对测试感兴趣，并想把它应用到课堂中，那么这本书很适合你。如果你想提高测试技巧，并研究出新的测试方式，那么这本书也很适合你。但是，如果你想充分了解有关测试的理论知识，那么这本书恐怕不是你的最佳选择。

《英语课堂教学评价活动》为中学教师编写，其中所提到的具体做法适用于中学生，但也可以通过简单的改变，用于高年级学生。不过测试儿童却是另外一回事，不包括在本书的内容中。

“我对测试一无所知，也不明白其中的术语。”

《英语课堂教学评价活动》首先讨论评估些什么，怎样评估。同时介绍测试中的术语，如，测试（testing）和教学（teaching）、评价（evaluation）和评估（assessment）、测量（measuring）和评判（judging）这几组词语之间有什么不同。（阅读第一章~第三章）

“如何判断测试的好坏？”

评估的传统方式就是使用试卷。语言测试是一个专业学术领域，包括大量的文献资料、理论背景、数据和语言测试方面的学术期刊。本书没有足够的篇幅对测试所涉及的内容面面俱到，所以我们主要介绍测试的基础知识：如何设计一份好试卷（阅读第四章）。同时我们也可以了解到，无论对于教师还是学校，要想设计一份好试卷并非易事。

“测试和评价有很多不同的形式吗？”

传统的测试观和现代评价观之间最大的区别在于作为教师和专业人员，我们不再重视死记硬背，而是转向帮助学生学会学习：包括学生的学习过程和学习态度。评价比单纯的测试包括的范围要广。第五章主要讨论测试和评价的形式。

即使在像外语这样的学科里，也有很多学习能力和技能需要学生掌握。那么我们如何把这些技能放在一起评估？哪些技能更重要？哪些技能可以忽略？我们可以拿教师办公室里的情况来打个比方。你不妨试一试这个极端的游戏：想象所有的教师都在你的办公室里面，你首先把所有的椅子围成一个半圆以开始讨论评价。然后宣布让所有的教师以他们能力的大小来就座，能力最“强”的老师坐在一端的椅

子上，能力最“差”的老师坐在另一端的椅子上。

当然，这一定会引起混乱！如何评估谁最好、谁最差呢？至少有关评估的所有问题都会引发争论。比如，有的教师在学科教学上很优秀，但在论文写作上却不很在行；有的教师很有想法，却不擅长管理班级等等。难道好想法比班级管理和学科教学更有价值吗？

“这些都是理论，
如何把测试理论
运用于实践呢？”

为了评估学习者及其学习过程，我们需要收集学生的数据：

- 学生能否运用语言的组成成分——语法、词汇和语音？（阅读第六章）
- 学生能否在阅读、写作、听力和口语方面运用语言本身？（阅读第七章~第九章）
- 学生如何最有效地学习？这是否可以成为测试的一部分？（阅读第十章）

“我应该在学年末
测试学生，还是一
学年测试两次，
或是更多？”

如果教师对学生的发展很感兴趣，则需要每隔一段时间就要评估一次。第十一章和第十二章讨论持续性评估（**continuous assessment**）和形成性评价（**formative evaluation**），即如何记录学生在语言学习和学习技能方面的进步。

通常在一门课程结束时，教师需要进行评估：学生的优势和弱势各是什么？过去，教师和学生都会退出评估，而让考试来做决定。实际上双方都没有真正地承担起责任。

“如果测试是一种
个体活动，那么
我要如何评估班
上的30人？”

教师一个人不可能承担起评估者的所有责任。但是有31个人可以来帮助每一个学生：教师、该学生和全班同学们。只要我们重新审视责任的分配，评估大班是有可能的。如果教师把所有的事情都帮学生做好了，学生干嘛还要费心去做呢？就拿作文来说，通常我们可能要花30分钟批改学生用10分钟所写的作文，而我们所发现的每一处错误都是学生没有发现的，或者他们根本懒得发现这些错误。

“应该如何使学生
更主动地投入到
评估过程中来？”

教师应该号召学生主动参与评估过程：不管我们多么努力，我们不能代替他们学习！第十三章主要讨论怎样做到这一点。共同承担责任意味着教师和学生都需要尝试新东西。革新是一件很有挑战性的事情。这意味着我们要学习一些新技能，同时会担心自己不能做得像以前那样好。

“应该如何使用
本书？”

把测试作为解决问题的方式

以下引言选自Arthur Hughes的《外语教师测试手册》（*Testing for Language Teachers*）（第6页）。他认为语言测试不是我们通过书面或口语说明如何去做，而是把它作为需要解决的问题。

语言测试者有时会被问及什么是最好的测试或最好的测试技巧。这种问题表明了人们对于语言测试的误解。实际上，没有最好的测试也没有最好的测试技巧。因此，可以得出结论：每一种测试情况都是独特的，从而产生特殊的测试问题。测试者的任务就是提出最好的解决办法来解决那个问题。

以上观点适用于所有评价：没有所谓“正确的”答案。只是提出了一个问题，而你们学校教师和学生的答案和同一条街上的另外一所学校的答案肯定是不同的。本书试图解释一些概念并提出一些观点。但不可避免的是，你可能会觉得书中的观点都很“理想化”，描绘了一个“完美的情境”。能用什么、如何使用以及可以忽略哪些你认为在你的情境中根本不可能使用的内容，这完全由你决定。和所有影响许多人的问题一样，任何解决办法都要通过讨论、争论后达成一致。这需要花一些时间，但我们应该让学生参与。因为我们的评估也许会长期影响他们未来的生活。

第一部分 评估、测试和评价

第一章

为什么评估学生的学习

许多群体对于评估学生的能力都很感兴趣：教师、教研组长、家长、政府。当然，还有学生自己。但是，我们开展评估主要有以下四点原因：

- 比较学生。
- 考查学生是否达标。
- 帮助学生学习。
- 检测课程是否有效。

任务

列举你在学校里参加过的各种考试（不仅是外语）。为什么举行这些考试？每次考试主要针对哪些群体而举行？结果提供给谁：学生、教师、教研组长、家长、政府还是其他？

1 比较学生

如果你的学生想进入一所大学学习热门专业，这所大学就必须选择要录取哪些学生。这就需要通过比较作出决定，例如，这所大学想要录取考生的前20%。但这就存在一个一致性（consistency）的问题。整体考得好的那年的考生可能要和整体考得差的那年的考生进行比较：今年的前20%不一定和去年的前20%一样好。但仍然是前20%的考生会被录取。这种方法叫做定量分配（rationing the carrots）：不论所有考生表现如何，只有前20%可以被录取。

尽管这种体系看似不公平，但还是常被政府和家长采纳来评判一所学校的质量。

2 考查学生是否达标

政府或国际考试委员会这些大型机构对于学生要达到的水平都有一定的标准。这些标准不一定完全反映学生所学的内容：不同的学校使用不同的教材或大纲。因此，这些大型机构必须确定自己的衡量准则，以考查学生是否达到标准。

其他小型机构，如个别学校，也可以自己确定标准。

但是，学校通常都会依据自己的教学内容进行测试。他们分析学生在课堂上学了些什么，然后通常通过学业成绩测试（achievement

test) 考查学生是否学会了。

测试者对于学业成绩测试所应测试的内容有不同意见。学业成绩测试可能用来测试:

- 大纲 (syllabus) 所规定的总体教学目标 (例如, 用英语表达过去时态的能力或用不同文体写作的能力)。
- 大纲中的具体项目 (例如, 一般过去时或撰写广告)。

评估的另一个原因就是为入学编班。我们可以通过分析学生的能力来考查其在某一群体中所处的位置。例如, 如果学校对于班级的规模有所限制, 就需要根据招收比例来编班 (例如, 前10%进入快班)。或者, 学生可能需要达到一定的标准。如果一个是写作班, 一个是语法复习班, 就要根据学生是否达到相应的标准进行分班。

3 帮助学生学习

无论是评估学生的语言水平还是学业成绩, 我们都可以用诊断的方式来分析学生的能力。除了通过评估来给学生打分, 我们还可以用诊断的方式考查学生在哪些方面需要更多的帮助。例如, 学生在撰写广告方面得分很高, 但在语法方面却犯了很多错误, 尤其是一般现在时的第三人称单数的使用方面。这样我们就可以在这方面给学生提供额外的帮助。

4 检测课程是否有效

假使所有学生在撰写广告方面都得了高分, 但在一般现在时第三人称单数的使用上都犯了很多错误, 我们就可以调整整个教学计划, 在这方面给全体学生提供额外的帮助。

从更大的范围来说, 如果教师 and 教学督导在所有的学校都发现同样的问题, 那么政府就会考虑改变整个教学计划。

小结

正如本书所介绍的, 评估学生有多种方式, 但最常用的评估方式可能就是测试。

- 水平测试 (proficiency test) 不考虑教学的具体内容, 而是衡量学习者的语言总体水平。
- 学业成绩测试 (achievement test) 是用来考查学生对所学内容的掌握情况, 也可以测试大纲上的具体项目或者总体目标。
- 编班测试 (placement test) 是上述两种测试的综合, 取决于我们采用什么样的标准来给学生编班。
- 诊断性测试 (diagnostic test) 用水平测试或学业成绩测试来分析学生或教学计划的优势和弱势。

任务

列举两种你很熟悉的测试：语言测试或你们学校所进行过的其他测试，另外一个科目或能力（如驾驶）的测试。

这种测试是否基于一定的教学计划？

谁来确定测试的标准？

测试结果用来做什么，比较学生、评估教学方案？还是其他原因？

第二章

测试、教学和评价

什么是测试？

每次我们让学生回答问题，而问题的答案是我们已知的，这时我们就在对学生进行测试。实际上我们在课堂上所做的大部分活动都是在测试学生的知识。例如：

He goes to the cinema. They...?

Find a word in the text that means 'angry'.

On the tape, where does John tell Susan he wants to visit?

What is the main idea of paragraph three?

Dictation: write down the following...

That's that part of the lesson finished. What do you think we're going to do next?

测试与教学

量化学生课堂表现

从传统意义上来说，测试用来衡量学生的课堂表现。

- 我们选取有代表性的语言点。
- 我们衡量学生是否可以运用这些语言点。
- 我们通过分数的形式来量化学生的能力。
- 我们记录下这些成绩，以此做出最终的评估。

一直以来，所有的测试理论（无论是语言测试还是洗发水的测试）通常都是基于准科学的程序，即：

1. 对行为进行衡量。
2. 采取措施来影响行为。
3. 再次对行为进行衡量，然后比较不同。

采取这种传统的模式对语言学习者进行测试意味着把语言学习者作为一种植物来对待。我们先来测试植物，然后采用新的肥料，之后再次测试植物，看看肥料所起到的作用。作为语言教学者，我们首先采用编班测验，进行教学，然后再采用学业成绩测试来检测学生的提高程度。

换句话说，测试通常与量化相联系，也就是把课堂表现转化成数字的形式。

	植物	语言学习者
阶段1	测试植物	测试一般现在时
阶段2	添加肥料	教授一般现在时
阶段3	再次测试 比较差异	再次测试一般现在时 比较差异

测试活动与教学 活动

测试和教学密不可分。我们经常通过提问题来检查学生是否理解所讲的内容。同样，提问也可以帮助我们发现需要教学的内容。我们本能地知道为什么要提问：无论是为了教学，还是为了测试。

比较下列两个练习：

Exercise 1

Fill the gap with an appropriate form of the verb.

- a John _____ France every year since 1993. (visit)
- b John _____ France last year. (visit)

Exercise 2

In groups, discuss the differences between the two sentences.

- a John has visited France every year since 1993.
- b John visited France last year.

练习1假设学生具备这方面的知识，并让学生通过做题来证明。很明显这是一个测试活动，值得注意的是即使学生做对了，我们也不知道他们为什么做对了。或许是猜对的，或许是凭感觉。

练习2就语言本身进行提问。换句话说，这种练习是让学生从中总结出可以运用在其他情境中的规律，也就是概括规则。同时也试图让学生意识到语言是怎样运作的。这是在帮助学生学习，是教学活动。当然，一些老师可能会认为，学生不需要知道为什么这样组织句子就是正确的，他们只要能用对就可以了。

让我们再来比较另外两个例子：

Exercise 3

Composition: A Summer's Day at the Beach (150 words)

Exercise 4

Read the following two compositions entitled 'A Summer's Day at the Beach'.

Which do you prefer and why?

Underline all the words and ideas relating to summer. Underline all the words and ideas relating to the beach. Put a tick next to the parts you like in each essay. Put a cross next to the parts you don't like in each essay.

If all the paragraphs got accidentally jumbled up, could you put them back in the right order? What would help you do this? Discuss your ideas with another group.

Homework: write down your own composition on the same theme (150 words).

跟前面两个练习的分析方法一样，练习3很明显就是一个测试活动，想让学生来展示他们能够做到什么程度。练习4则试图让学生更清楚地意识到他们能够做的事情：在给出任务之前首先让学生增强意识。这样可以帮助学生学习。

教学还是测试？

有时，即使是教师也弄不清楚他们究竟是在教学还是在测试。我们可能认为自己是教学生，而实际上却在考学生。

这一点在我们教听、说、读、写四项技能时体现得特别明显。此时，语言教师面临着一个主要的问题：我们自己对这些技能知道的也不够多，对于好的听力、好的阅读或好的其他技能并没有明确的尺度。我们所知道的就是一些归纳性的概念，如略读和跳读，但这些并不够具体，无法帮助我们设计出有效的、帮助学生不断提高的课程。

换言之，在面对一项很难教的技能，如好的听力技能，我们通常以两种方式中的一种来解决这个难题。其一，我们为学生提供大量的机会来展示他们所知道的，以此来检测他们是否提高了。我们让学生读、写或听不同难度的文章，同时希望学生可以保持在同一程度上甚至有所提高，或者我们保持文章不变，但增加问题的难度。

这有点像医生说：“我不知道你生病的原因，也不知道你为什么好转，但是你的体温确实降下来了。”在教这四项技能时，我们所能做的就是给学生提供语言素材，并通过测试来衡量他们的“体温”，以此检测他们是否提高了。

其二，我们用容易教的技能来代替难教的技能。尽管有些技能的规则不是很清楚，但语法和词汇是有明确规则的（但是，设计语法和词汇的测试是很复杂的，这点我们在以后的章节中会详细讨论），我们更容易接受这些规则。因此，有时我们会以为自己在教授或测试一项技能，而实际上却在练习或测试语法或词汇知识。例如，许多口语测试实际上是在复习语法，即通过口头的形式来测试语法。这类测试无法真正测试口语能力，如口语能力之一：会话中如何插话而不会冒犯对方。

为什么会产生这样的情况？因为我们先前所提到的准科学的植物测试程序存在一些问题。下面我们主要讨论这些问题。

测试中的问题

问题1：技能的量化

在第7页，我们看到测试是基于科学的一种理念：测量—改变—再测量—比较。

这一科学模式存在的问题就是并非所有的东西都能够通过这种方式来测量。有些知识用这种方式来测量很容易，如，一般现在时的第三人称单数。

但也有一些技能很难衡量。例如，我们如何量化学生对班级做出有用贡献的大小？

- 首先，我们必须以某种可以测量的方式来定义“有用”和“贡献”。
- 我们可以定义“有用”为“成功地向其他学生解释某事”。
- 我们可以定义“贡献”为“成功回答了老师向全班同学提出的问题”。
- 现在我们可以统计一个学生有多少次成功地回答了老师提出的问题，而且班上多数学生都理解了。

问题在于，现在我们只是计算出了一个学生有多少次成功地回答了老师提出的问题，并且班上多数学生也能够理解。但这与一个学生对班级做出有用贡献未必是一回事。

因此在评估一些较难的技能时就存在两个危险：

- 我们可以对都理解的事物重新定义，使其可以测量；但这样做的同时，可能改变了我们要测量的事物本身。
- 如果有些技能很难测量，我们就不放进测试题，即使这些技能很重要。

最终结果就是，我们只是在测量容易测量的事物，而不是真正去评估我们想提升的学生的实际表现。

任务

注意听在办公室里教师之间使用第一语言的对话。

在他们的交谈中完整的句子占多大比例？

有多少是由一些“啊”、“呀”等语气词所连接的句子片断？

在你的实际教学中有多少时间是在教学生一些句子片断呢？

测试中的其他问题

问题2：结果还是过程，是什么还是为什么

定量测量（quantitative measurement）这一准科学方法的另一个问题就是它不能记录定性资料（qualitative data）。测量可以告诉我们植物是否生长了，但不能说明其为什么生长。它提供给我们有关结果的信息，但对于过程却无法提供任何帮助。

通过第8页的练习4，我们可以对学生的能力有一个全面的了解，因为我们可以看到过程，如，让学生在文章中打对号或打叉，然后看看这些活动的效果能否或如何反映到学生自己的作文中去。

问题3：标准化和偶然结果

这一准科学方法存在的第三个问题就是给植物施加的肥料必须是一样的，否则结果就很难比较。我们必须排除变量，以便来评估课程的实施是否成功。但在教学中做到这一点是很难的。这意味着在所有的学校中教学内容必须完全一致，否则我们就无法对个别学生的成长过程进行

比较。这种测试的模式最终导致的是以教师为核心的权威式教学法。

这种科学方法更侧重于一般的发展规律，而忽略个体的特殊性。假设在一次听力测试中，所有其他的学生都得到分数的90%，而最优秀的学生却只得到10%。作为教师，正是这个出乎意料的结果值得我们去研究。

任务

任选一本教科书，可以是你在学校中所用的，随机选择三个听力练习、三个阅读练习和三个口语练习。每个练习的目标是什么：

- 测试语法还是词汇？（如，Mr Brown _____ the cinema等）
- 测试学生的理解力？（如，通过多项选择题检测阅读理解；信息沟活动等）
- 教会学生听、读或说得更好？（如，是否包括如何改进阅读、听力、练习插话等建议）
- 教会学生学习？（如，是否教课堂用语或能否帮助学生回答自己提出的问题等）

测试和评价

测试和评价的关系就如同课程（curriculum）和教学大纲（syllabus）的关系一样。

教学大纲是教师在一个学期内需要完成的具体教学项目，但教学大纲又是更大的方法论体系的一部分，即课程。语言教学不仅要教些什么，还包括如何去教（即课堂教学过程），以及为什么去教（即大纲或课堂教学过程背后的教学方法或理论基础）。

评价

在本书中，评价所包括的范围比测试要广。评价把测试作为一项有用的工具，同时也包括其他用来评估学生行为的重要方法。我们想要评估学生使用一般现在时的能力，同时也想获得他们语言学习的其他信息，如：

- 他们是否会用词典？
- 他们是否在课堂上使用目标语言？（例如在对话时）
- 他们的笔记是否记得很有条理？
- 他们是否参与小组讨论？
- 他们的课堂表现是否良好？

如果我们把评价、测试与教学大纲、课程用图表相比较，就可以看到它们的相似性：