

致 谢

这本书从一个想法发展到现在您手中所拿到的成书，我们得到了许多人的意见、建议和帮助。在此对下列人士谨致谢忱：

感谢《语言测评实践：现实世界中的测试开发与使用论证》一书的作者为本书提供了整体的概念性框架。我们对这一框架进行了调整和拓展，使其适应课堂语言测评的具体场景。

感谢以下同仁在本书撰写过程中对各章节、各版本提供的反馈，他们是 Nathan Carr、辜向东、何莲珍、Lorena Llosa、Gary Ockey、Adrian Palmer、Yasuyo Sawaki 及其学生、Masaki Eguchi、Kana Matsumura、Tatsuro Tahara、Takumi Uchihara。

感谢以下同仁在他们的语言测评课上试用了本书的部分章节，他们是 Nathan Carr（美国加州州立大学富勒顿分校现代语言文学系 TESOL 项目）、辜向东（中国重庆，重庆大学外国语言与文化学院英语系；学生负责人：钟瑜）、何莲珍（中国杭州，浙江大学外国语学院）、Gary Ockey（美国爱荷华州立大学英语系）和 Yasuyo Sawaki（日本东京，早稻田大学教育学院英语语言文学系）。

感谢我们在世界各地举办的工作坊的参与者，我们在工作坊中介绍了基于“测评使用论证”的课堂语言测评方法，参与者们对这些方法的反馈、提问和建议都非常重要，帮助我们了解了一线语言教师的需求和“真实情景”，使我们的课堂语言测评方法最终得以成形。

感谢牛津大学出版社的 Alex Birtles 和 Antoinette Meehan 对我们的信任，相信我们能够为教师提供对学生进行课堂语言测评的新方法。

感谢 Antoinette 在本书的编辑和出版过程中给予的帮助，感谢 Helen Wendholt 和 Danielle King 对本书终稿的编辑。

各位匿名审稿人对本书的最初方案给出了建设性意见和建议，在此一并感谢。

Lyle Bachman

美国加州大学洛杉矶分校

Barbara Damböck

德国迪林根教师培训学院

2017年8月

目 录

| | |
|----|--|
| 致谢 | |
|----|--|

第一部分：引言

| | |
|---------------|---|
| 第一章 课堂语言测评新方法 | 3 |
|---------------|---|

| | |
|----------------|---|
| 1.1 方法源于何处? | 4 |
| 1.2 本书因何而写? | 5 |
| 1.3 本书为谁而写? | 6 |
| 1.4 本书包括哪些内容? | 7 |
| 1.5 本书能让你学到什么? | 8 |
| 1.6 参考文献 | 8 |

| | |
|--------------|---|
| 第二章 使用课堂语言测评 | 9 |
|--------------|---|

| | |
|--------------------|----|
| 2.1 语言教学与课堂语言测评 | 9 |
| 2.1.1 语言教学 | 9 |
| 2.1.2 语言测评 | 10 |
| 2.1.3 语言教学与语言测评的异同 | 11 |
| 2.2 基于课堂测评的决策 | 11 |
| 2.2.1 课堂上需要做出的决策 | 12 |
| 2.2.2 决策的相对重要性 | 14 |
| 2.3 课堂测评模式 | 17 |
| 2.4 使用课堂语言测评 | 20 |

| | |
|---------------------|----|
| 2.5 测评研发者与使用者 | 25 |
| 2.6 测评使用的可行性 | 26 |

第二部分：课堂测评的系统观

| | |
|--------------------|----|
| 第三章 测评研发方法概述 | 31 |
|--------------------|----|

| | |
|-------------------------|----|
| 3.1 公平性与问责制 | 31 |
| 3.2 测评合理性论证 | 32 |
| 3.3 测评研发 | 35 |
| 3.4 为测评使用论证提供支撑证据 | 38 |

| | |
|----------------------------|----|
| 第四章 后效：为什么需要对学生进行测评? | 40 |
|----------------------------|----|

| | |
|--------------------------------|----|
| 4.1 主张 1 的测评研发问题 | 41 |
| 4.1.1 预期后效 | 41 |
| 4.1.2 谁会受到后效的影响? | 43 |
| 4.1.3 预期后效的特征 | 44 |
| 4.1.4 小结：课堂测评的预期后效和利益相关者 | 45 |
| 4.2 陈述主张 1 | 46 |
| 4.3 为主张 1 提供支撑证据 | 49 |

| | |
|---------------------------|----|
| 第五章 决策：何时需要对学生进行测评? | 52 |
|---------------------------|----|

| | |
|--------------------------------|----|
| 5.1 主张 2 的测评研发问题 | 52 |
| 5.1.1 需要做出什么决策? | 53 |
| 5.1.2 谁将受到预期决策的影响? | 53 |
| 5.1.3 谁将做出预期决策? | 54 |
| 5.1.4 何时需要做出预期决策? | 54 |
| 5.1.5 预期决策的特征 | 54 |
| 5.1.6 小结：在课堂测评中关联预期后效与决策 | 56 |
| 5.2 陈述主张 2 | 56 |

| | |
|------------------------------------|-----------|
| 5.3 为主张 2 提供支撑证据 | 60 |
| 第六章 解释：测什么与怎么测？ | 63 |
| 6.1 主张 3 的测评研发问题 | 63 |
| 6.2 测什么？预期解释 | 65 |
| 6.2.1 确定做决策所需的解释：相关性和充分性 | 65 |
| 6.2.2 能力定义的来源 | 66 |
| 6.2.3 定义所要测评的语言能力范围：有意义 | 68 |
| 6.3 怎么测？测评任务 | 69 |
| 6.3.1 解释的概推性 | 70 |
| 6.3.2 确定目标语使用域和目标语使用任务 | 70 |
| 6.3.3 解释的公正性 | 71 |
| 6.3.4 解释的特征 | 71 |
| 6.4 关联主张 3 与主张 1 和 2 | 72 |
| 6.5 陈述主张 3 | 73 |
| 6.6 为主张 3 提供支撑证据 | 75 |
| 第七章 测评记录：如何记录学生的测评表现？ | 78 |
| 7.1 主张 4 的测评研发问题 | 79 |
| 7.2 测评记录 | 80 |
| 7.2.1 确定我们需要的测评记录类型 | 80 |
| 7.2.2 测评记录的特征 | 81 |
| 7.3 陈述主张 4 | 83 |
| 7.4 为主张 4 提供支撑证据 | 89 |
| 第三部分：在课堂上应用这种方法 | |
| 第八章 研制测评任务模板 | 95 |
| 8.1 步骤 1：选择目标语使用任务 | 97 |

| | |
|---|------------|
| 8.2 步骤 2: 描述目标语使用任务 | 98 |
| 8.3 步骤 3: 修改目标语使用任务, 研制测评任务模板 | 104 |
| 8.4 步骤 4: 研制记录 / 评定学生表现的方法 | 114 |
| 8.5 步骤 5: 从测评特征角度对修改后的任务 / 测评任务模板 进行评价 | 115 |
| 第九章 创设语言测评任务 | 119 |
| 9.1 用模板创设多项测评任务 | 119 |
| 9.2 研发考查相同能力的多项测评任务示例 | 120 |
| 9.2.1 示例 9.1: 描述性段落写作 | 121 |
| 9.2.2 示例 9.2: 讨论环境问题 | 125 |
| 9.2.3 示例 9.3: 听力中理解细节信息 | 129 |
| 9.2.4 示例 9.4: 阅读中理解衔接 | 134 |
| 第十章 组合测评任务研发课堂测评 | 139 |
| 10.1 什么是蓝图? | 140 |
| 10.2 如何用蓝图指导测评研发? | 146 |
| 10.3 如何用蓝图进行质量控制? | 146 |
| 10.3.1 比较测评任务与教学任务 | 147 |
| 10.3.2 比较不同测评中的任务 | 147 |
| 第十一章 施测程序和指令 | 148 |
| 11.1 为什么要明确施测程序? | 148 |
| 11.2 施测程序 | 149 |
| 11.2.1 准备测评场景 | 149 |
| 11.2.2 提供轻松的测评环境 | 150 |
| 11.2.3 发布指令 | 151 |
| 11.2.4 何时给出指令? | 151 |
| 11.2.5 如何让指令可理解? | 153 |

| | |
|--|------------|
| 11.2.6 指令需要多详细? | 154 |
| 第十二章 评定学生的测评表现和报告结果 | 156 |
| 12.1 对单个测评任务或题目进行评分..... | 157 |
| 12.2 不同类型的评分方法..... | 158 |
| 12.2.1 按成功完成的任务数进行评分 | 159 |
| 12.2.2 按语言能力水平进行评分 | 161 |
| 12.3 计算总分和报告结果..... | 163 |
| 12.3.1 单项测评的总分 | 164 |
| 12.3.2 基于多项测评的分数 | 168 |
| 12.3.3 报告学生学业情况的方法 | 169 |
| 12.4 参考文献..... | 171 |
| | |
| 第四部分：课堂语言测评示例 | |
| 测评示例 | 175 |
| 示例 1：口语：描述动物（少儿，初级） | 178 |
| 示例 2：阅读：格林威治小镇，句子填空（青少年或成人， 高级） | 189 |
| 示例 3：口语：描述我最喜欢的地方（青少年，中级） | 199 |
| 示例 4：听力：回答关于业余活动的问题（青少年，中级） | 207 |
| 示例 5：听力：回答关于生命中最美好的时刻的问题（青少 年，初级） | 218 |
| 示例 6：写作：描述乐器（青少年，中级） | 231 |
| 示例 7：口语：讨论环境问题（青少年或成人，高级） | 240 |
| 示例 8：写作：求职信（青少年或成人，高级） | 249 |
| 示例 9：阅读：标题匹配，街头艺术（青少年，中级） | 259 |
| | |
| 附录 1 使用语言测评前需要思考的问题清单 | 269 |

| | |
|---------------------------|-----|
| 附录 2 测评任务模板的组成部分 | 272 |
| 附录 3 第二章和第三章活动的参考答案 | 276 |
| 术语表 | 280 |
| 延伸阅读建议 | 286 |
| 索引 | 288 |

第一部分

引言

第一部分介绍本书并概述课堂测评。第一章介绍课堂测评方法的起源、撰写本书的初衷、本书的目标读者及主要内容。第二章讨论测评在教学中的作用，介绍课堂测评相关的概念，概述使用课堂测评的过程。

第一章

课堂语言测评新方法

在考虑测评学生语言能力时，你会问自己哪些问题？何时以及以何种频率测评？测评学生语言能力的哪些方面？使用何种测评任务或题目？用多少测评任务或题目？如何评分？分值怎么定？如何确保测评结果一致并能提供关于学生语言能力的有意义的信息？

以上是老师们经常问到的有关课堂测评的问题。每一个问题都是你在计划测评学生语言能力时需要考虑并应对的，这些问题又涉及不同的测评议题。不过，这些议题虽各不相同，但却并非相互独立，而是相互关联的。要回答使用什么样的测评任务这一问题，你需要知道要考查的是学生语言能力的哪些方面。同理，要回答何时或以何种频率测评这一问题，你需要知道用这些测评信息做什么。对这些单一问题的回答既不足以形成一种针对学生测评的综合性方法，也无法帮你研发出一些能够向学生、家长/监护人、学校管理者或其他对测评结果感兴趣的人证明其合理性的测评。最重要的是，也无法让你研发出令自己信心满满的测评项目。

本书将呈现一种对学生进行测评的全新方法。运用这一方法，你首先要问自己一个问题：“我想通过测评带来哪些有益后效？”例如，“测评将如何帮助我提升自己的教学水平？”“测评将如何帮助我的学生学得更好？”在这一方法的指引下，你就能把预期的有益后效与“考什么”（语言能力的某些方面）和“如何考”（测评任务的类型）联

系起来。遵循这一过程，你就能研发并使用课堂测评，这些测评与课堂教学紧密关联，并支撑教学。也就是说，测评将为教学活动的组织以及学生的学习提供信息。

1.1 方法源于何处？

我们的课堂语言测评方法是基于现有的理论与实践总结的。在过去大约 25 年的时间里，语言测评领域发生了巨大的、令人兴奋的变化，具体体现在语言测评观念的形成以及语言测评的研发与使用上。语言测评不再被视为技术主导型活动，即获得考试分数并对分数进行统计分析，而是被广义地视为一种嵌入了社会文化因素的过程。具体的语言测评也不仅仅被视为测量工具，更被视为反映某种文化价值观的活动，并通过测评实现这些价值。就其研发和使用而言，语言测评主要关注的不再是考查什么能力、使用什么测评类型，而是测评对于个人和机构而言会有什么影响。最后，语言测评领域的研究范围也在扩大，不仅仅关注涉及重大决策的大规模测试，也开始关注能帮助老师改进教学、帮助学生提升学习效果的课堂测评。表 1.1 总结了这些进展。

表 1.1 语言测评进展

| | 过去 | 现在 |
|-------------|------------------------------|------------------------------|
| 语言测评 | 被视为技术主导型活动，即获得考试分数并对分数进行统计分析 | 被广义地视为一种嵌入了社会文化因素的过程 |
| 具体的语言测评 | 被视为测量工具 | 被视为反映某种文化价值观的活动，并通过测评实现这些价值 |
| 作为研究领域的语言测评 | 只关注涉及重大决策的大规模测试 | 拓展了研究范围，包括能为改进教学与学习提供参考的课堂测评 |

Bachman (1990) 和 Bachman & Palmer (1996, 2010) 的著作不仅为上述进展做出了贡献, 同时也从中汲取了养分。我们的课堂测评方法正是基于《语言测评实践: 现实世界中的测试开发与使用论证》(以下简称《语言测评实践》)(Bachman & Palmer, 2010) 一书中适用于一般场景的语言测评方法提出的。Bachman & Palmer 的语言测评方法被测评研发者用来指导高风险、大规模的语言测评研发和使用。与本书更密切相关的是, 这一方法也被广泛应用于教师的职前培训课程(如文凭、证书、硕士学位、博士学位等), 并在世界范围内被应用于许多面向一线教师的在职语言测评培训课程和工作坊。

本书中讨论的许多概念、术语和理念均源于 Bachman & Palmer 的《语言测评实践》一书。但由于本书并非纯学术理论著作, 因此在提及《语言测评实践》的内容时, 我们并没有采用每次提及均引用的方式, 谨在此做统一声明。

1.2 本书因何而写?

我们在对语言教师的职前培训和在职培训过程中积累了丰富的经验。在过去的 30 多年里, Lyle 在美国高校教授本科生和研究生的语言测试课程, 并在世界各地为一线教师举办过多个工作坊。Barbara 有 20 年的中小学语言课程教授经历, 并在德国担任了 8 年的教师培训师。在过去 5 年中, 我们携手为一线教师举办了多个工作坊。

通过多年来对一线语言教师的观察以及与他们合作的经验, 我们熟悉了他们对课堂测评的需求、对课堂测评的了解和看法, 以及对自身测评能力的焦虑和关切。我们了解到, 总体来说, 大多数语言教师对语言测评的了解和看法都是基于自身经验, 而非有关语言测评方面的教育或培训。我们还了解到, 那些受过语言测评专业培训的 language 教师所参加的课程或工作坊主要针对大规模测试研制的设计原则和测试技术, 而这些原则和技术可能与课堂测评并不完全相关。这类课程往往聚焦语言测试的基本原则, 如信度、效度、真实性以及测什么和怎么测的技术问题(如考查语言能力的哪些方面、命制什么样的题目、

如何对学生的作答进行评分、如何对分数进行统计分析等)。

根据经验,我们认为有必要研制一种不同的语言测评方法,以满足一线教师的需求。这种方法聚焦研发与使用有助于改进教学和学习的语言测评,可赋能教师,使之更有效、更有信心地对学生的语言能力进行测评。在研制这一方法时,我们对 Bachman & Palmer (2010)一书中提出的整体思路进行了拓展与调整,并将其应用于课堂语言测评中。

在教师培训工作坊中使用这一方法时,我们发现教师认为该方法直观易懂,并且与他们课堂测评的需求和关切紧密相关。我们关于如何更有效地在课堂测评中使用这一方法的思考也在一个接一个的工作坊中得到不断完善。由于该方法得到了工作坊中一线教师的积极评价和不断的完善,我们希望与更多的人分享该方法。最好的分享方式就是写一本书,也就是你手中的这一本。

本书的写作是一个不断演进的过程。根据牛津大学出版社的编辑和匿名审稿人提出的意见和建议,我们对最初方案进行了修改。当然,我们也希望得到更多其他读者的反馈,因此邀请了一些熟悉的、教授职前培训课程的测评专家在课上试用本书各章节的内容。非常荣幸的是,我们邀请的每一位专家(详见致谢名单)都欣然答应,并提供了宝贵的反馈意见。

1.3 本书为谁而写?

如果你是一名一线语言教师,想知道如何更好地测评学生的语言能力,本书就是为你量身定制的。书中含有大量示例和活动,适合自学。本书专为一线语言教师而写,包括:

- 1) 各类学校(公立或私立)、各个层次(学前教育、小学、中学、大学、成人教育)、语言培训学校或培训项目的语言教师;
- 2) 在不同国家教授具有不同文化和语言背景的学生的语言教师;
- 3) 使用不同教学方法(教学法、教学大纲、教材)的语言教师。

如果你正在参加职前培训课程或接受在职培训,本书也同样适合

你。基于收到的反馈，我们认为，本书适合从事下列活动的学生和教师群体：

- 1) 修读语言教学证书或文凭课程；
- 2) 攻读应用语言学、语言教学、语言测评硕士学位；
- 3) 参加在职培训项目。

1.4 本书包括哪些内容？

本书有以下几个特点，旨在激发你的兴趣，包括：

1) 对概念和测评程序的讨论

在讨论部分，我们介绍并讲解新的材料。这些材料大多相互关联并具累积性，因此我们会将新材料与已介绍过的材料进行关联。这一循环讨论过程会导致一些概念重复出现，但我们认为这有助于巩固你所积累的知识。

2) 术语的定义

我们相信书中对专业术语的定义能帮助你理解和记忆。书中的术语均用粗体，并用箭头标识，便于识别，如下所示：

➤ **语言使用任务**是要求个体使用语言实现某一特定目标的活动。

这些术语在书后的术语表中按字母顺序列出。

3) 例证

大多数章节都有一些示例，用于阐明讨论中提及的要点。这些示例选自本书第四部分所列的完整测评示例。

4) 活动

大多数章节包括活动板块，我们希望这些活动能够加深你对讨论中提及的概念的了解。这些活动的目的不是要测试你对概念的理解，而是要引导你对讨论中提出的具体问题思考，并给出可能的回答。活动答案详见附录 3 以及第四部分的示例。

5) 示例

除上述特点以外，本书第四部分呈现了完整的语言测评任务示例。这些示例说明了不同年龄段（少儿、青少年、成人）、不同语言能

力水平（初级、中级、高级）群体的四类语言活动（听、说、读、写）的测评任务。你可以轻松地用这些示例所提供的模板自主研发测评任务。这些示例为你结合自己的课堂测评情景设计测评任务提供了可复制的任务样例。这些示例经过改编，可用于其他语言使用活动、其他年龄段或其他能力水平的学生的测评（更多示例详见网站 www.oup.com/elt/teacher/lact，2023年12月15日读取）。

6) 其他特点

- (1) 附录 1：一张可供快速查阅课堂测评步骤的清单
- (2) 附录 2：测评任务模板组成部分汇总
- (3) 附录 3：第二章和第三章活动的参考答案
- (4) 延伸阅读建议

1.5 本书能让你学到什么？

我们相信，使用我们的方法，你将学会研发、使用课堂语言测评所需的知识和技能。这一方法将助力你改进课堂测评使用，帮助你自信地向学生、家长 / 监护人、学校管理者以及其他对你的测评结果感兴趣或想使用你的测评结果的人证明测评的合理性。

祝君阅读愉快，测评快乐！

1.6 参考文献

- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (2010). *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World*. Oxford: Oxford University Press.

第二章

使用课堂语言测评

本章讨论一些概念和问题，我们认为老师需要先熟悉这些概念和问题，然后才能理解测评在语言课堂中的作用和使用。具体而言，首先，我们讨论语言教学以及语言测评在语言课堂中的作用，并介绍两种不同的课堂测评模式；然后，我们讨论基于课堂测评所做的决策；接下来，我们讨论使用课堂语言测评的过程，以及测评研发者和使用者在使用语言测评过程中所发挥的作用；最后，我们探讨语言测评的可行性，以及确保测评切实可行所必须考虑的因素。

2.1 语言教学与课堂语言测评

如果想理解课堂语言测评，你需要置身于语言课堂这一语境。换言之，你需要思考语言测评如何作为语言教学的一部分来发挥作用。你可以尝试提出如下问题：

- 1) 什么是语言教学？为什么要教？
- 2) 什么是语言测评？为什么要测？
- 3) 语言教学和语言测评有何异同？

2.1.1 语言教学

在语言课堂上进行教学时，你会让学生参与到语言使用任务或需要使用语言的任务中来。这些语言教学/学习任务的目的是帮助学生学习语言。也就是说，你通过让学生参与语言教学/学习任务来帮助

他们提高语言能力。

- **语言使用任务**是要求个体使用语言实现某一特定目标的活动。
- **语言教学**指让学生参与以提高语言能力为目的的语言使用任务的过程。
- **语言教学 / 学习任务**是老师和学生共同参与的**语言使用任务**，其目的是提高学生的语言能力。

语言教学和语言教学 / 学习任务的主要目的是改进学生的语言学习。然而，这些任务还有其他目的，如促进学习者认知、情感和社交等方面的发展，激励或帮助学生培养对所学语言和文化的积极态度。

2.1.2 语言测评

在语言课堂上对学生进行测评时，你也会向学生呈现需要他们使用语言的任务。而这些语言测评任务的目的是收集学生的语言能力信息——他们学了什么、学了多少。

- **语言测评任务**是一种语言使用任务，其目的是采集学生的语言使用样本。（注：语言测试人员通常使用“题目”这个术语，其含义与我们这里所说的“测评任务”基本相同）。
- **语言测评**这个词有两种用法。其单数形式（a language assessment）指的是许多语言测评任务或题目的集合。作为总括性概念（language assessment），它指收集学生语言使用样本的过程。
- **课堂语言测评**是由一名或多名教师研发和 / 或在课堂上使用的语言测评。

需要指出的是，语言测试人员可能会使用不同的术语，如“测评”（assessment）、“测量”（measurement）、“测试”（test）等，来指一项本质相同的活动——收集信息。一些语言测试人员明确区分了测评和测试。但在本书中，我们对这两个术语基本不做区分。

把语言测评看作是一个过程这一点很重要，因为对学生进行语言测评只是整个测评过程的第一步，后续还需要根据测评结果对学生的语言能力进行解释，并基于该解释做出相应的决策，为学生、老师、学校或其他个人和机构带来有益后效。

2.1.3 语言教学与语言测评的异同

教学和测评都是需要师生共同参与的过程或活动。教学和测评过程中，老师都会向学生呈现需要他们使用语言的任务。教学和测评的主要区别在于它们的目的不同。教学旨在促学，而测评旨在收集信息。表 2.1 总结了教学与测评之间的异同。

表 2.1 教学与测评的异同

| | 教学 | 测评 |
|------|------------------------------------|------------------------------------|
| 参与者 | 老师和学生 | 老师和学生 |
| 任务要求 | 老师向学生呈现学习任务； 学生参与到需要他们使用语言的任务中来 | 老师向学生呈现测评任务； 学生参与到需要他们使用语言的任务中来 |
| 任务目的 | 促进学生学习； 促进学习者在语言、认知、情感和社交方面的发展 | 收集学生语言使用表现的样本 |

虽然语言教学和语言测评目的不同，但它们互为参考。语言教学以两种方式为课堂语言测评提供信息。第一，对所测评的语言能力的定义应基于语言教学的内容或学习目标——教学大纲、教材和学生要参与的教学任务。第二，所使用的测评任务往往是基于课堂上使用的教学/学习任务。同样的，测评也可以为教学提供信息。通过测评获得的信息既可以帮助你调整教学方式，也可以帮助你需要对需要复习或强调的教学内容进行调整。

2.2 基于课堂测评的决策

通过语言测评获得的有关学生语言能力的信息可用于决策。教师可以用这些信息做出其认为既有助于改进教学又能帮助学生改进学习的决策。教师也可以根据课堂测评的结果给学生评定等级或级别。在许多国家，其他人也会参考这些等级或级别来对学生做出决策。

2.2.1 课堂上需要做出的决策

在课堂上，决策始终是教学的一部分。你需要对使用什么材料、给学生什么样的学习任务、什么时候安排复习以及什么时候开始上新课等问题做出决策。有时你会决定调整教学材料或教学任务，或决定给学生提供反馈，以帮助他们进行自主学习。你可能需要对学生进行分组或分级，或决定哪些学生将升入高一级别或开始新课的学习。

通过测评收集到的有关学生语言能力的信息可以帮助你做出上述决策。例如，如果你在一节课结束时给学生做了小测验，发现大部分学生没有达到学习目标，你或许会决定再次使用这些教学材料，或者改变教学材料的使用方式，或者改变学习任务。如果你需要决定哪些学生课程学习合格或可以进入高一级别或年级，你很可能在课程教学过程中通过若干次测评来确定哪些学生已经达到了课程学习目标。当然，你也可以用这些测评为学生提供学习反馈。

决策的目的

教师基于课堂测评所做的决策可以根据决策目的进行分类。以下两个术语有助于你思考并记住你所做决策的不同目的。

➤ **形成性决策**是旨在改进教学和促进学习的决策。这类决策：

- 会涉及教学任务、教学材料或教学大纲的调整；
- 会涉及向学生提供反馈；
- 会涉及对学生进行分级，以促进教学和学习；
- 可在教学和学习之前、之中或之后做出。

➤ **终结性决策**包含两类决策：

- 升级：确保进入高一级别或下一门课程学习的学生能从新起点的教学中受益，留在本级别的学生在继续接受同一级别的教学时，也能够从中受益。
- 认证：确保通过某一语言能力水平认证的学生达到了该认证所规定的语言能力水平。

终结性决策：1) 涉及对学生进行分组，如升级组或不升级组，获得证书组和未获得证书组；2) 是在教学和学习过程结束后做出的决策。

需要注意的是，形成性决策和终结性决策都可能在教学和学习的过程结束之后做出。例如，你可能会在完成一个单元的教学后给学生做一次测试。根据测试结果，你可以向学生提供反馈，或对教学进行调整（形成性决策）。你也可以把这次测试的结果作为整个课程成绩的一部分，用以决定学生是否合格（终结性决策）。表 2.2 总结了形成性决策和终结性决策之间的区别。

表 2.2 形成性决策与终结性决策的区别

| 形成性决策 | 终结性决策 |
|----------------------------|---|
| 旨在改进教学和学习。 | 旨在确保学生能从适合其水平的教学中受益，或确保学生具备相应能力后才能获得认证。 |
| 涉及教学和 / 或学习任务、教材以及教学大纲的调整。 | 涉及对学生进行分组。 |
| 决策可以在教学和学习之前、之中或之后做出。 | 决策在教学和学习过程完成后做出。 |

活动 2.1 下列决策中哪些是形成性决策，哪些是终结性决策？在下表第二列或第三列对应的方框内打钩（√）。这些决策在什么时候做出——教学和学习之前、之中还是之后？在第四列中写出做出每项决策的时间：之前、之中或之后（第一个答案已经给出）。

| 决策 | 形成性决策 | 终结性决策 | 决策时间 |
|-------------------------|-------|-------|-------|
| 1. 调整教学（如教材、教学任务） | √ | | 之中或之后 |
| 2. 调整学习方法或学习策略 | | | |
| 3. 把学生分到语言学习项目中合适的组别或级别 | | | |

（待续）

(续表)

| 决策 | 形成性决策 | 终结性决策 | 决策时间 |
|------------------------|-------|-------|------|
| 4. 确定哪些学生可以升级 | | | |
| 5. 确定哪些学生可以修读高一级别的语言课程 | | | |
| 6. 确定学习重点 | | | |
| 7. 确定是否开始新的教学单元 | | | |
| 8. 认证学生的语言能力水平 | | | |
| 9. 确定教学重点 | | | |
| 10. 确定要补充的教学材料 | | | |
| 11. 确定哪些学生课程学习不合格 | | | |

可影印 © 牛津大学出版社

(活动答案见第 276—277 页)

2.2.2 决策的相对重要性

其他人也可能参考课堂测评结果对学生做出决策。你通常会用等级或级别表示学生在一个教学过程结束后的学业情况，这个过程可能是一个教学单元、一周、一个月、一个学期或一个学年。成绩评定通常是基于教学过程中学生参加的多次课堂测评结果。你给学生评定的成绩，尤其是一学年结束后的成绩，可能会成为大学录取的部分依据，也可能成为用人单位决定是否录用的依据。

教师和其他人根据课堂测评做出的决策都可以根据其相对重要性来加以描述。基于课堂测评所做的决策对于利益相关者来说并非同等重要。

- ▶ **利益相关者**是指可能受测评使用影响或得益于测评使用的人、语言项目或课程，或者机构。我们自己（教师）和我们的学生（考生）始终是利益相关者。此外，就某一项特定的测评而言，利益相关者可能还包括我们的同事、学校管理者、学生家长/监护人、雇主、政府官员和公众。

有些决策会对利益相关者产生重大影响。例如，学生离校时应具备哪些资格这一决策会对他们的人生产生重大影响，至少会在一定程度上决定哪些学生能被中学或大学录取，或者哪些学生能找到理想的工作，还可能影响到教师、学校以及学生的家长/监护人。有些决策则并不那么重要。例如，决定给某个学生反馈，告诉他需要花更多时间练习发音，这一决策就不太可能对他的人生产生重大影响。

以下三个术语可以用来描述所做决策的重要性。这些术语能够帮助教师更好地讨论基于测评所做决策的相对重要性。

- ▶ **低风险决策**是次要决策。如果决策错误，对于利益相关者而言成本很低，纠错也容易。决定向学生提供课堂测评表现反馈，调整教学任务等，都是低风险决策的例子。
- ▶ **高风险决策**是重大的、影响人生的决策。一旦决策错误，对于利益相关者而言成本很高，纠错很难。决定学生能否通过某门课程，哪些学生能获得毕业资格，或哪些学生能被大学录取等，都是高风险决策的例子。

所做决策的重要性是相对的，风险程度从低到高不等。如上面的例子所示，许多决策的风险很低或很高。然而，很多基于课堂测评所做的决策，既不是高风险决策，也不是低风险决策，这就需要用第三个术语“中风险决策”来描述了。

- ▶ **中风险决策**是风险不是很高也不是很低的决策。把学生分到某个语言项目的不同级别，决定哪些学生能进入某个课堂的高一级别等，都是中风险决策的例子。

你需要了解所做决策的重要程度，因为这将决定你在研发测评时投入多少时间和精力。决策越重要，需要投入的时间和精力就越多。