

目 录

第一章 形成性评价概述 (Formative Assessment: An Introduction)	1
第一节 形成性评价的定义	1
第二节 形成性评价的分类	23
第三节 形成性评价的特点	30
第四节 形成性评价的工作模型	37
第二章 教师主导的形成性评价 (Formative Assessment by Teachers)	53
第一节 课堂提问	53
第二节 形成性评价信息的反馈	75
第三节 学生学习档案袋	89
第四节 教师主导的课堂提问、反馈及学习档案袋	102
第三章 学生主导的形成性评价 (Formative Assessment by Students)	104
第一节 自我评价	104
第二节 同伴评价	115
第三节 同伴辅导	128
第四节 自我调节	132
附 录	141
第四章 形成性评价与终结性评价 (Formative Assessment and Summative Assessment)	153
第一节 形成性评价与终结性评价的关系	153
第二节 形成性评价和终结性评价对学习的影响	165
第三节 形成性评价与终结性评价的实践融合	170
小 结	191

第五章 形成性评价案例 (Cases of Formative Assessment)	193
第一节 临海	193
第二节 深圳	201
第三节 北京	204
第四节 写作档案袋	219
小 结	225
附 录	227
第六章 国内二语课堂形成性评价的现状与未来 (The Current Situation and the Future of Formative Assessment in L2 Classrooms in China)	239
第一节 国内二语课堂形成性评价研究与实践现状	239
第二节 形成性评价研究与实践中的主要问题	256
第三节 形成性评价未来的发展方向	267
第四节 汉语作为第二语言的形成性评价研究	280
结语 逃离分数的洞穴 (Closing Marks)	289
参考文献	291
致 谢	331
后 记	332

第一章 形成性评价概述

(Formative Assessment: An Introduction)

评价历来是教育体系中不可或缺的一部分。教育系统中的评价一般分为两大类：终结性评价和形成性评价。大规模、高利害的终结性评价是推进学校改革和进步的主要推动力之一，一直被视为影响学校政策和课堂实践的重要因素。在中国的外语教育中，有很多终结性、高利害的评价，其中教师较为熟悉的是中考、高考、大学英语四、六级考试等。随着教育改革的推动，与教师日常教学关系更为密切的形成性评价取得了一些进展，但仍处于初始阶段。要使形成性评价切实得以实施，教师对其充分的认识和了解是十分必要的。本章主要从形成性评价的定义、分类、特点和工作模型四个方面展开论述。

第一节 形成性评价的定义

(Definition of Formative Assessment)

形成性评价已有四十多年的历史，它对学生学习和教师教学的促进作用已经得到了研究者、实践者和政策制定者的一致认可。许多国家和地区不仅将形成性评价写进了权威的教育文件，而且将其视为评价体系中不可或缺的部分。尽管如此，众人关于形成性评价的理解仍然莫衷一是，有许多模糊甚至偏颇。据 2007 年 9 月 17 日《教育周刊》(Bennett, 2011) 报道，测试行业在形成性评价方面的分歧如此之大，以至于测试专家 Stiggins 都已经不再使用形成性评价这一术语 (Cech, 2007)。这种概念理解上的差异和不足，不仅使理论研究难以形成合力，也导致了实践推广上的举步维艰。具体来说，如果我们不能有效地定义形成性评

价，那么，(1) 我们就不能记录它的有效性，不能评价形成性评价是否已按计划实施，也不能将其推广到更多的课堂、学校和地区；(2) 我们就不清楚形成性评价研究总结报告中究竟应该涵盖哪些内容，因而无法对研究结果进行总结；(3) 我们就不能将其运用到我们的课堂环境中去，因为我们不了解在运用过程中要关注哪些要点 (Bennett, 2011)；(4) 将会阻碍教师对自己的教学实践进行总结，教师无法进一步提炼和扩展已经包含形成性评价的教学实践，不利于形成性评价在课堂层面的发展创新。

实际上，形成性评价因其促进学习的作用引起了许多学者的注意。他们给了形成性评价诸多定义。这些定义既有相同之处，又有互补和矛盾的地方。这种多样性一方面说明了形成性评价研究本身的复杂性，另一方面也确实为与其相关的研究和实践增加了难度。如果研究者和实践者对于何为形成性评价的理解都不一致，那么研究和实践便失去了理论基础。因此本节将对形成性评价的各种定义进行回顾和分析，比较其特点，旨在构建一个综合、全面、合理的形成性评价概念。

一、形成性评价定义的共性

首先，我们从形成性评价的各种定义所强调的五个方面来看其共性：目的、学习成就证据、评价任务、时间和频率以及反馈。

1. 强调形成性的目的

多数学者都从目的和使用的角度来定义形成性评价。首次使用“形成性”这一术语的学者 Scriven (1967) 给出了如下定义：

当教育者对一个成熟的、已经完成的教育项目的价值进行评价时，他所进行的是终结性评价，其目的是评判，帮助决策者做出关于教育项目是否合格的决定；但当教育者对一个正在进行中的教育项目进行评价时，他所进行的则是形成性评价，其目的不是评判，而是修正，即找出项目中的潜在问题，并做出如何改进项目的决定。

从以上定义来看，Scriven 当时提出的形成性评价与我们现在所理解的主要在课堂中进行的形成性评价有所不同，它的评价对象不是学生的学习情况，而是某个教育项目的质量。继 Scriven 之后，“终结性”和“形成性”这两种评价方式被 Bloom 引入教学评价领域。Bloom (1968)

认为形成性评价是在教学过程中的每一个阶段提供反馈和纠正措施，是教师和学生通过小测试进行的评价，其目的是辅助教学过程。Bloom, Hastings 和 Madaus (1971) 合作出版了《学生学习形成性评价和终结性评价手册》，重新定义了形成性评价与终结性评价，为课堂教学形成性评价的发展奠定了基础。他们将“终结性评价”这一术语定义为在学期、课程或项目结束后所进行的评价，其目的是为了打分、评级、评价进展，或进行课程、研究项目、教育计划的有效性研究，最核心的特征就是在学习或教学完成之后对学习或教学效果做出判断。但是这种判断常常引起学生、教师和课程开发者的焦虑和防御心理。在现实生活中，虽然我们不希望进行这种形式的评价，但又不得不使用它。而形成性评价是在课程建设、教和学的过程中发生的各种系统性评价，旨在改进课程及教与学。在形成性评价中，我们关注课程建设、教与学中的产出，不仅将这些产出作为最有价值的关于课程以及教与学的证据，而且还寻求最有效的方法来报告这些证据，如通过降低与评价相关的负面效果，或让教师、学生以及课程开发者（形成性评价的使用者）做出判断。

之后的二十多年中，形成性评价并没有得到足够的重视，直到 Black 和 Wiliam (1998a) 开始对形成性评价进行综述研究，才又掀起了形成性评价研究和实践的热潮。Black 和 Wiliam 继承了 Bloom 从目的或功能的角度定义形成性评价的传统，将其定义为所有教师 / 学生进行的活动，而且这些活动提供的信息将用于反馈，以调整随后进行的教与学的活动。

此外，众多学者都给出了各自对形成性评价的定义，大多数都强调了评价的功能。在这些定义中出现了“to advance”“to modify”“to improve”“to enhance”等表示改进之意的词语。

2. 强调以学习成就证据为基础的调整

从目的和使用的角度来看，任何课堂活动或任务都可以是形成性的，只要该活动或任务产生的信息被用于指导和调整教和学，以促进学习 (Black & Wiliam, 1998a)。但这种调整不是随意发生的，形成性评价强调的对教和学的调整必须以学生的学习成就证据为基础。

Black 和 Wiliam (2003) 认为“任何评价活动，只要它提供的信息被教师和学生用作反馈，以调整他们教和学的活动，那么这种评价就有

利于学习。当学习证据被用于调整教学工作以满足学习需求时，评价即为形成性评价。”同样，Heritage 等（2009）将形成性评价定义为：在教学过程中连续收集学习证据、提供学习反馈的系统化的过程。在这一精心计划的过程中，教师或学生使用以评价为基础的证据来调整正在进行的教和学的活动（Popham, 2008）。

在 Popham（2006）看来，基于学习证据的调整或改进是形成性评价中不可或缺的因素，是形成性评价的“本质”。Popham 认为形成性评价的本质在于它以评价引出的学生学习状态的证据为基础，教师调整教学策略、学生调整学习策略。教师和学生的判断可能并不总是正确，但基于证据的调整优于毫无根据的调整。这种基于证据的调整看似简单的“目的一手段”模式，不值得加以注意和研究，但是许多人类行为都是基于这一模式，人类的诸多进步都以此模式为基础。通过分析过去行为的结果来指导将来行为的人，无疑将获得更大的成功，教师的教和学生的学也不例外。

3. 强调评价任务的开放性

如上文所述，形成性评价强调教和学的调整应该以学习证据为基础，因此，产生学习证据的评价程序是形成性评价必不可少的一部分。与以纸笔测试为主要形式的终结性评价相比，形成性评价产生学习证据的方式各异。当然，形成性评价也可以采用纸笔测试的方式获取学习证据，不过它更多地采用非传统的、非正式的方式获取所需的学习证据。Irons（2008）认为任何使学生学习产生了反馈（或前馈）的任务或活动都可以称作形成性评价。Black 和 Wiliam（2003）认为，只要设计和实践的首要目的是为了促进学习的评价，都可以称为形成性评价。至于通过何种方式获得学习证据，则并无任何限制。Black 和 Wiliam（1998a）的定义更是直接指出了教师和学生的活动都具有形成性的价值，即形成性评价是一个开放的体系，它没有限定的方式或策略。学生的口头或书面表达、肢体动作、检查表、同伴对话、传统测验、概念图、日志、评价标准、苏格拉底式提问都可以为形成性评价提供信息。只要能让我们看到学生的认知过程，从表现性评价到多项选择题都可以成为形成性评价的形式。重点是上述活动可以让学生展示自己的思维过程，方便教师引出关于思维过程的证据。

实施形成性评价并不是一件简单的事情，并不是在现有的实践基础上冠以“形成性评价”之名就能很快产生效果。形成性评价在课堂中的表现形式多样，每个教师、每种教学风格都可以找到适合自己的形成性评价的形式。因此，教师需要做的不是将某种现有的实践形式拿来使用，而是要真正理解形成性评价的精神，根据教学和学生的需求，找到适当的方式，再将形成性评价的特点融入到自己的课堂实践之中。

4. 强调评价时间和频率

时间维度是定义形成性评价时需要关注的另一要点，因此大多数定义都强调形成性评价是在教学过程中进行的。比如，Cowie 和 Bell (1996) 将形成性评价定义为：在学习过程中，教师和学生用于了解学生的学习状况，并对此做出回应以促进学生学习的过程。Wininger 和 Norman (2005) 认为，形成性评价就是在教学开始前或进行过程中，为了调整教学并提升学生表现，对学生的学习进度进行测量。Heritage 等 (2009) 将形成性评价定义为：在教学过程中连续收集学习证据、提供学习反馈的系统化过程。美国教育考试服务中心 (ETS) 更是明确地将形成性评价描述为：“教师和学生使用学习证据调整教和学以满足学生每时每刻 (minute by minute) 和每天 (day by day) 的即时学习需求”。

从上述定义中可以看出，形成性评价是在教学过程中发生的，甚至是每时每刻都在发生的评价活动。这与 Bloom 最初对形成性评价的理解有些不同。在 Bloom 的概念中，形成性评价是在 1—2 周的单元教学后进行的形成性测试。从不同时期人们关于形成性评价的定义中我们发现，形成性评价发生的频率越来越高。这一方面表明，将时间和频率作为形成性评价的关键特征已经不太合适了，另一方面也表明形成性评价与教学密切结合的趋势。正如 Sadler (1989) 所说，形成性评价与终结性评价的区别在于目的和效果，而不是评价的时间。

5. 强调反馈的使用和效果

Sadler (1989) 认为，反馈是形成性评价中的一个关键因素。但他强调，只有在被用于缩短学生现有的理解水平和理想的理解水平之间的差距时，反馈才能体现其形成性功能。而这一点是由学生控制的，即只有当反馈给学生的信息被学生用于提升学习表现时，它才是形成性的。

换言之，形成性评价的形成性不在于有没有反馈，而在于反馈有没有被使用。更有极端的观点认为，应该从产生的效果去定义形成性评价。Wiliam (2000) 认为：意图为形成性（有形成性目的）的评价，但是最终没有达到预想效果（没有实现形成性功能）的评价不能称为形成性评价。Popham (2006) 和 Shepard (2008) 也持相同的观点，强调除非教学得到改进，产生形成性的效果，否则不能使用形成性评价这一术语或称其为形成性评价。

我们认为，从有没有提高学习的效果这个角度去定义形成性评价或许欠妥，因为影响学生学习效果的因素复杂，反馈只是其中一个因素，要想做到准确地归因非常困难。其次，教或学效果的验证可能需要相应的方法和长时间跟踪，在可行性上存在问题。因此，较为合理的判断方法是评价有没有对教师和学生的行为产生影响。产生影响的依据在于教师和学生是否使用反馈，如果是，就说明反馈对教与学产生了影响。当然最终效果的体现方式、效果大小、效果好坏，也不是形成性反馈这一个因素就可以解释的。Kluger 和 DeNisi (1996) 在研究中指出，他们所回顾的大部分研究的不足之处在于这些研究只考虑到了短期的效果。他们建议，关于反馈的研究更应该关注反馈所引发的学生对不同类型反馈的反应或处理方式，而不是反馈对学生的知识技能表现产生的影响。因此，应该从反馈有没有得到教师和 / 或学生的使用、对他们的行为有没有产生影响和改变的角度去定义形成性评价，而教师和学生因评价而产生的行为改变有没有提升知识和技能可能是另一个研究问题。

二、形成性评价定义的分歧

随着人们对形成性评价的认识不断深化，不少国家的教育评价推广组织和众多学者纷纷给出了形成性评价较为全面的定义。但是到目前为止，学界对形成性评价的本质是什么还存在不同的看法。

不过，如果因为对“形成性”这个词的不同理解而放弃“形成性”这个概念，也许并非明智之举。首先，这个术语已被广泛使用，而且具有相当大的影响力，我们不可能要求人们在短时间内放弃这一概念。其次，“形成性”一词体现了教育的核心，反映了教育的主要目标是帮助学习者发展和成长。在我们看来，关于形成性评价理解的种种分歧，一方面体现了形成性评价丰富的发展历程，另一方面也体现了审视形成性评价的不同视角。了解这些分歧，有利于我们更好地理解和把握形成性评价。

目前，关于形成性评价定义的分歧集中在以下三个方面：形成性评价的主导者、形成性评价的计划性以及形成性评价的过程性。

1. 形成性评价：教师主导与学生主导

从现有的形成性评价的文献来看，大家对形成性评价认识上的分歧之一在于形成性评价究竟是由教师主导、针对教，还是由学生主导、针对学。许多对形成性评价的定义都从教的角度出发，强调评价对教学的影响。如 Learning Point Associates (2009) 认为形成性评价是一个过程，在这个过程中教师使用各种工具和策略判断学生的知识水平，发现差距并规划未来教学以改进学习。Gareis (2007) 认为，形成性评价指任何用于教师发现学生学会了哪些内容、没有学会哪些内容的手段，其目的在于教与学，而不是评分。Stiggins (2002) 指出促学评价要求教师使用课堂评价过程以及评价过程所提供的信息，以促进学生的学习，而不仅是检查学生的学习状况。Tunstall 和 Gipps (1996) 认为，形成性评价是指教师使用对于学生知识或理解状况的判断，为教学过程提供反馈，决定教师是否需要重新解释任务或概念，进行进一步的练习，或直接进入下一个阶段。Harlen 和 James (1996) 尝试发展形成性评价的概念，旨在帮助教师有效使用教学过程中收集的学生信息。

以上所有定义都强调了教师的作用，认为评价控制权掌握在教师手中，是教师使用评价信息，做出关于下一步教与学的决定；学生在收集、阐释评价信息和接下来的教师干预中处于静态。换言之，学生在形成性评价过程中似乎是被动的，知识的学习和掌握是教师干预调整教学的结果。

当然，一些强调教师主动性的研究者在他们发展形成性评价概念的过程中也提到了学生的作用，如 Harlen (1996) 认为，形成性评价本质上是为教师和学生提供学生目前理解和技能发展状况的信息，以决定将来的前进方向。但是，学生在评价过程中的具体作用并不明确。如果形成性评价的目的是为了促进学习，那么根据社会文化理论和建构主义学习理论，必须强调学生在理解、阐释、内化评价过程以及反馈中的作用。一些研究者注意到了这个问题，在形成性评价的概念化过程中，明确了形成性评价过程由学生和教师共同主导，强调了形成性评价是教师和 / 或学生使用的过程 (Popham 2008; Perie 等, 2007; CCSSO, 2008c;)。William (2011) 更明确地指出，在形成性评价的过程中，教

师、学生或同伴都要参与学习信息的收集、阐释和使用，突出了学生参与形成性评价的过程。Wiliam 定义的不足之处在于，他将这些过程的目的限于帮助做出关于下一步教学的决定，而没有提及学生学习策略的调整。实际上，在形成性评价的过程中，学生和教师的作用同等重要，他们在互动合作中共同参与形成性评价的全部过程。

2. 形成性评价：计划与非计划

Popham (2008) 强调了形成性评价的计划性。形成性评价包括一系列经过认真考量的、由教师和 / 或学生开展的活动。在 Popham 看来，教师在课堂上进行即时的现场教学调整固然很好，但不能算作形成性评价，因为如果要达到研究证据证明的效果，形成性评价必须是认真计划的过程。Popham 强调计划性的积极意义在于，他指出了形成性评价虽然发生在教学过程当中，但在具体教学评价之前就应启动相关工作，教师教学前的设计、思考和准备工作也十分重要。此外，这种对计划性的强调，还出于对评价目标的考虑。形成性评价关注的重点应该是促进高层次认知技能的发展，所以实施过程必须经过详细周密的计划和安排 (Popham, 2011)。这种看法确实具有积极的作用，但它的缺陷也同样明显。Popham 对计划性的强调致使形成性评价的范围缩小，将强调相倚性的课堂即时互动评价排除在形成性评价的范围之外。

课堂中的形成性评价应该是教师和学生之间能够对学习产生积极影响的社会互动 (Torrance & Pryor, 1998)。如果从目的和功能的角度来定义形成性评价，课堂即时评价无疑能起到很好的促学作用 (杨华, 2012)，特别是对于二语课堂学习来说。非计划的形成性评价能够提供更为丰富、全面的学生信息，如学生的个人发展与社交能力发展的信息。而且，对于二语教学这样的课程来说，非计划的互动性形成性评价所提供的关于学生技能发展的信息有时更即时、更可靠、更深刻，也更能促进学生语言能力的发展。因此，至少对于二语课堂来说，将互动性形成性评价排除在形成性评价之外的做法欠妥。除此之外，即便是课堂即时评价，也并非没有计划性的成分。即时的教学调整需要教师具有良好的知识储备。学生对教师计划好的教学内容不理解，是教学中常常发生的状况，而教师对此的回应也应该是教师事前计划的一部分。将课堂即时评价作为形成性评价的一部分，在一定程度上对教师形成性评价的